

Bayesian Transductive Markov Random Fields for Interactive Segmentation in Retinal Disorders

Noah Lee, Andrew F. Laine
Heffner Biomedical Imaging Lab (HBIL)
Department of Biomedical Engineering
Columbia University, New York, NY USA
{nl2168, laine}@columbia.edu

R. Theodore Smith
Department of Ophthalmology
Columbia University, New York, NY USA
rts1@columbia.edu

Abstract - In the realm of computer aided diagnosis (CAD) interactive segmentation schemes have been well received by physicians, where the combination of human and machine intelligence can provide improved segmentation efficacy with minimal expert intervention [1-3]. Transductive learning (TL) or semi-supervised learning (SSL) is a suitable framework for learning-based interactive segmentation given the scarce label problem. In this paper we present extended work on Bayesian transduction and regularized conditional mixtures for interactive segmentation [3]. We present a Markov random field model integrating a semi-parametric conditional mixture model within a Bayesian transductive learning and inference setting. The model allows efficient learning and inference in a semi-supervised setting given only minimal approximate label information. Preliminary experimental results on multimodal images of retinal disorders such as drusen, geographic atrophy (GA), and choroidal neovascularisation (CNV) with exudates and sub-retinal fibrosis show promising segmentation performance.

Keywords: *Interactive Segmentation, Naive Bayes, Mixture Models, Generative Learning, Markov Random Fields, Bayesian Transduction.*

I. INTRODUCTION

Over the last decade we have seen an increase in the amount and complexity of heterogeneous biomedical data coming from laboratory tests, different imaging methodologies, and gene-protein analysis revealing information at nano-to-organ scales. Current clinical practice in ophthalmology requires the physician to extract, correlate and interpret information from heterogeneous and multimodal data sources over time to diagnose, treat and manage patients with retinal disorders [4]. Nowadays this is still a manual and time intensive process. Retinal imaging such as color fundus (CF), fluorescein angiography (FA), autofluorescence (AF), infrared (IR), red-free (RF), and optical coherence tomography (OCT) provides the physician with multimodal pathologic and -anatomic information enabling quantitative assessment of retinal disorders. Specific to the disease, the phenotype manifestations contain high variation in shape, topology, structure, appearance, and geometry as well as noise and artifact degradations. Automatic segmentation in these domains is still an unsolved problem [2].

In the realm of computer aided diagnosis (CAD) interactive segmentation schemes have been well received by

physicians, where the combination of human and machine intelligence can provide improved segmentation efficacy with minimal expert intervention [1-3]. Transductive learning (TL) or semi-supervised learning (SSL) are suitable frameworks for learning-based interactive segmentation given the scarce label problem [5, 6]. We aim to derive a segmentation model that adaptively learns from the data and the expert. In the medical domain the labeling process of multimodal image data requires expert knowledge and time intensive editing to obtain accurate label information for the object, which is to be quantified. Minimizing expert intervention time is a desired criterion. In this regard, transduction offers a workaround by leveraging the labels provided by the expert to label the remaining unlabeled test set [5]. Obtaining accurate label information of the object of interest enables quantification of pathologic information that can be used for exploring functional and temporal relationships to other indicative disease factors to assess the patient's health condition over time [7].

In this paper we present extended work on Bayesian transduction and regularized conditional mixtures for interactive segmentation [3]. The main contribution of the paper is twofold. First, we present a Markov random field model integrating a semi-parametric conditional mixture model within a Bayesian transductive learning and inference framework for spatial classification. A special case of our general framework is the transductive Gaussian mixture model or the optimal transductive Bayes classifier. The transductive generative formalism allows us to provide i) predictive confidence of the classification and ii) assess performance guarantees of the inference while exploiting correlation between labeled and unlabeled data. The model allows efficient learning and inference in a semi-supervised setting given only minimal approximate label information. Second, we report experimental results on the application of multimodal interactive segmentation of retinal disorders such as drusen, geographic atrophy (GA), and choroidal neovascularisation (CNV) with exudates and sub-retinal fibrosis. We present promising preliminary results (see fig. 2 right column) where segmentations are obtained with minimized expert intervention (see fig. 2 left column). In addition we compare the model with wr.t. spatial regularization constraints. The figure below shows examples of 4 different types of multimodal ophthalmologic imagery.

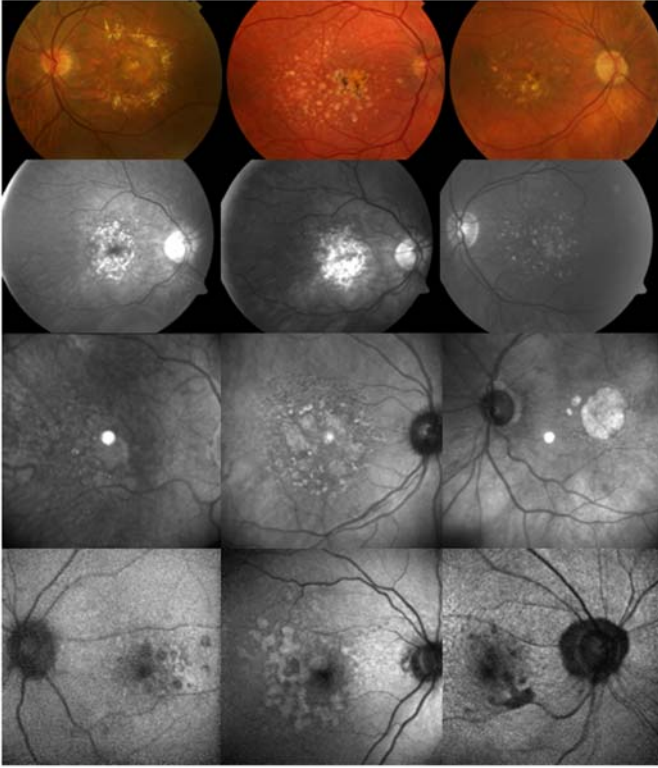


Fig. 1. Retinal Disorders in Multimodal Ophthalmologic Imagery. The top row shows color fundus images of CNV with exudates and drusen patients. The second row shows fluorescein angiography images of GA patients. The third row shows infrared images of GA (with central acquisition artifacts) and the bottom row shows autofluorescence images of GA. Shown images exemplify the segmentation problem.

II. RELATED WORK

Consider a dataset $\mathcal{D} = [\mathcal{D}_l, \mathcal{D}_u] \in \mathcal{P}_{XY}$, where $\mathcal{D}_l = \{(\mathbf{x}_n, y_n)\}_{n=1}^l$ denote the labeled training set and $\mathcal{D}_u = \{(\mathbf{x}_n, \hat{y}_n)\}_{n=l+1}^{l+u}$ the unlabeled test set with \hat{y}_n unknown. The usual case is $l \ll u$. We refer the reader to [3, 8, 9] for closely related work. Markov random fields (MRF) [8, 9] are a special case of undirected graphical models for modeling spatial dependencies in a probabilistic generative framework [10]. Conditional independence assumptions allow for local factorizations of the joint distribution

$$p(x) = \frac{1}{Z} \prod_{\mathbf{c}} \psi_{\mathbf{x}_C}(\mathbf{x}_C) = \frac{1}{Z} \exp\left\{-\sum_{\mathbf{c}} E_{\mathbf{x}_C}(\mathbf{x}_C)\right\}, \quad (1)$$

where Z is the partition function and $\psi_{\mathbf{x}_C}(\mathbf{x}_C) \geq 0$ are strictly positive arbitrary clique potentials over a clique \mathbf{x}_C . In section III we outline on how to transform $\psi_{\mathbf{x}_C}(\mathbf{x}_C)$ into a Bayesian transductive model for spatial regularized segmentation.

III. BAYESIAN TRANSDUCTIVE RANDOM FIELDS

Given the scarce label problem we choose a generative model over the discriminative counterpart motivated in part by a faster convergence rate of the asymptotic generalization error [11] when label information is scarce. In particular, we present a Markov random field model using conditional mixture naïve

Bayesian assumptions within a transductive learning and inference setting. Compared to [12] and [13] our model assumes for the class-conditional likelihood a Gaussian mixture for each covariate feature dimension allowing us to represent and describe more complex distributions while keeping the parameter space tractable. In addition the MRF constraint allows for spatial regularization, whereas [12, 13] applied their work to the application of natural language processing (NLP). To simplify the estimation we reduce the parameter space by assuming naïve conditional independence between the feature space and the class label imposing a regularization constraint on the class-conditional likelihood. A special case of our model is the transductive Gaussian mixture model or the optimal transductive Bayes classifier. The naïve conditional independence assumption allows efficient computation of marginal and conditional distributions [14] suitable for large scale learning and inference. The posterior is formed by learning the class-conditional mixture model $p(x|y)$ and prior $p(y)$ for each class exploiting labeled and unlabeled data. We allow the posterior distribution to depend on the unlabeled test set \mathcal{D}_u with spatial MRF regularization constraints to exploit the smoothness- and cluster assumption between the marginal $p(x)$ and the conditional $p(y|x)$ to perform predictive segmentation.

A. Transductive Learning and Inference

Our modeling problem consists of two latent variables one for $\mathcal{P}(y|x)$ and the other for approximating the marginal \mathcal{P}_X . Given \mathcal{D}_l we learn the class-conditional and unconditional mixture densities of each class by maximizing the log-likelihood of $p(x|y)$ and $p(y)$. To learn the marginal \mathcal{P}_X for a given class label we assume $p(x)$ to be distributed as a Gaussian mixture on each covariate feature dimension. To approximate both latent variables Z and Y we build the following likelihood model on \mathcal{D}_l and \mathcal{D}_u

$$\mathcal{P}(\mathcal{D}|\theta) = \prod_{n=l+1}^l p(y_n|\pi) \sum_{j=1}^d \alpha_j p(x_{j_n}|y_n, \theta) \times \prod_{n=l+1}^{l+u} p(y_n|\pi) \sum_{j=1}^d \alpha_j p(x_{j_n}|y_n, \theta). \quad (2)$$

The maximum a posteriori (MAP) estimate of parameter θ for $\mathcal{D} = [\mathcal{D}_l, \mathcal{D}_u] \in \mathcal{P}_{XY}$ with $l + u$ independent and identically distributed (i.i.d.) observations has no closed form solution. Taking the log-likelihood of (2) gives

$$\mathcal{L}(\hat{\theta}|\mathcal{D}) = \sum_{n=1}^l \log p(y_n|\pi) + \sum_{n=1}^l \log \sum_{j=1}^d \alpha_j p(x_{j_n}|y_n, \theta) + \sum_{n=l+1}^{l+u} \log \hat{p}(y_n|\hat{\pi}) + \sum_{n=l+1}^{l+u} \log \sum_{j=1}^d \alpha_j \hat{p}(x_{j_n}|y_n, \hat{\theta}) \quad (3)$$

The log-sum term of above log-likelihood in equation (3) is a marginal probability and requires a non-linear optimization

scheme. Alternatively equation (3) can be optimized by an iterative method to obtain the MAP or ML solution. One can choose from belief propagation and other approximate inference algorithms in probabilistic graphical models [ref]. We choose the EM algorithm [15] for the sake of simplicity and conceptual clarity. Lower bounding the log-sum term with an auxiliary function $\mathcal{L}(q, \theta)$ a local solution can be obtained by iteratively ascending

$$\begin{aligned} \text{E-Step} \quad q^{t+1} &= \arg \max_q \mathcal{L}(q, \theta^{(i)}) \\ \text{M-Step} \quad \theta^{(i+1)} &= \arg \max_{\theta} \mathcal{L}(q^{(i+1)}, \theta). \end{aligned} \quad (4)$$

A proof that the update equations in (4) indeed maximize the log-likelihood can be found in [15]. The ML estimate of the sum-log term of equation (3) is much simpler. Maximizing the log-likelihood with respect to π the solution to the constraint optimization problem for the labeled training data is:

$$\hat{\pi}_{ML} = \arg \max_{\pi} \sum_{n=1}^l \log p(y_n | \pi) = \sum_{n=1}^l y_n / l \quad (5)$$

Analog to equation (3) the ML estimate for π on the unlabeled test set updates accordingly with changed summation indices. From (4) and (5) the maximum a posteriori classification on the unlabeled test set can be obtained in a straight forward manner. Inference is performed given the unlabeled test set through the Iterated Conditional Modes (ICM) algorithm [9].

$$\hat{y}_i = \arg \max_{y_i \in K} P(y_i | y_N, x_i) \quad (6)$$

In the MRF model, assuming above likelihood and pairwise interactions among adjacent labels, one maximizes the local conditional probability assuming observation independence

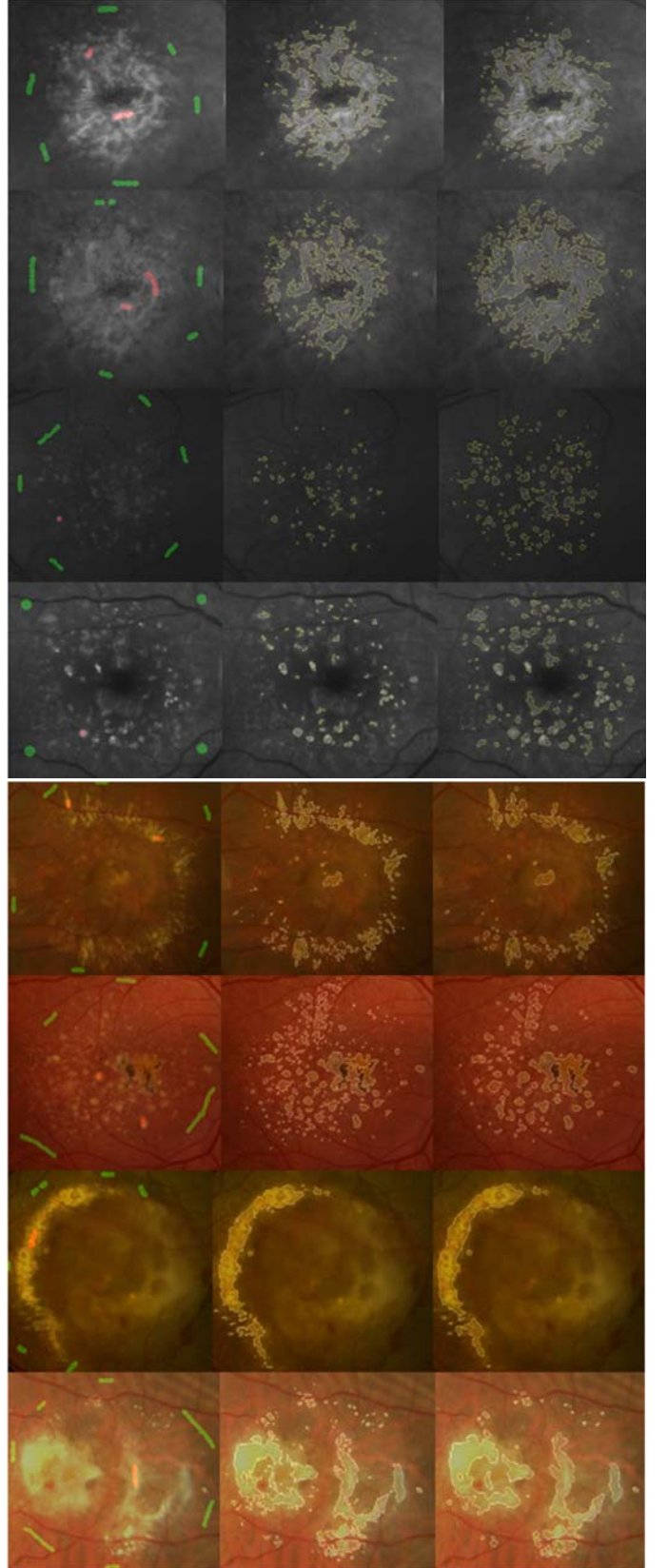
$$\hat{y}_i = \arg \max_{y_i \in K} \frac{1}{Z_i} \exp \left\{ \log(P(x_i | y_i) + \sum_{j \in N_i} \beta y_i y_j) \right\} \quad (7)$$

with β denoting a weighting parameter for spatial regularization, Z_i the partition function for each class i .

IV. EXPERIMENTS & RESULTS

We have applied our model on real world multimodal ophthalmologic image segmentation. We evaluated our model on 4 different modalities including CF, AF, IR, and RF. The expert provided approximate label information marking object (red) and background (green) regions for the object of interest. Model comparisons for two sets of parameters $\beta = \{0, 1.5\}$ are shown. Images from 16 patients have been centered around the macula and cropped to 324x324 regions to reduce image size and improve visualization. We could observe promising segmentation performance, where the expert continued the labeling process until the desired labeling was reached. In general the model with spatial regularization, as expected,

provided smoother label outlines and improved segmentation results confirmed by the expert.



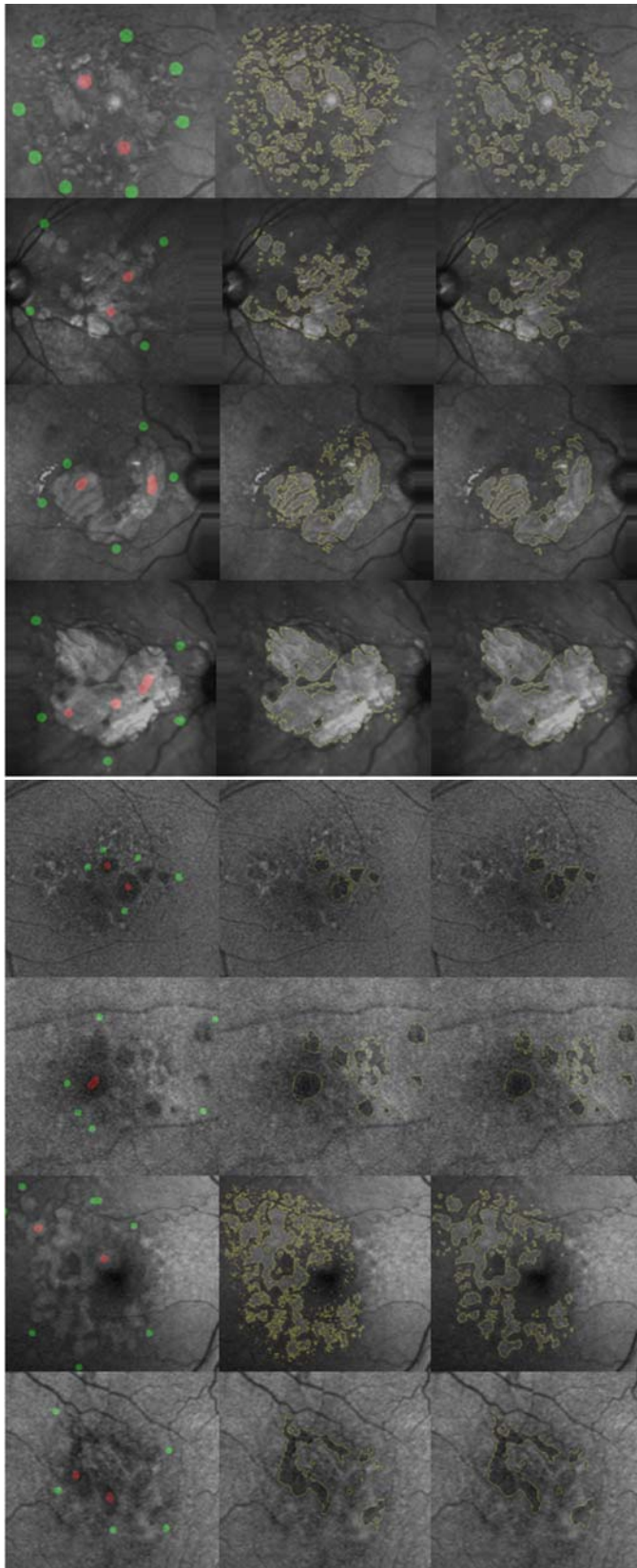


Fig. 2. Qualitative segmentation performance. (LEFT) Expert provided labels. (MIDDLE) Segmentations obtained without spatial regularization $\beta = 0$. (RIGHT) Segmentations obtained with spatial regularization $\beta = 1.5$. For

visualization purposes segmentation outlines are overlaid in yellow for gray scale images and white for color images. The color images show segmentation of exudates, drusen, and subretinal fibrosis respectively.

V. CONCLUSION

In this paper we have presented a novel model for spatial multimodal interactive segmentation. Bayesian transductive Markov random fields (abbreviated BTMRF) fuses concepts from probabilistic generative learning and spatially constrained Bayesian inference. The model allows efficient learning and inference in a semi-supervised setting given only minimal approximate label information. Preliminary results show promising performance of difficult and various segmentation complexities in multimodal ophthalmologic imagery.

VI. ACKNOWLEDGMENTS

This work was supported by the National Eye Institute (NEI) (R01 EY015520), the New York Community (RTS), and unrestricted funds from Research to Prevent Blindness (RPB).

REFERENCES

- [1] N. Lee, A. F. Laine, and R. T. Smith, *Coarse to Fine Segmentation of Stargardt Rings using an Expert Guided Dual Ellipse Model* 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), p. 177, Vancouver, Canada, (2008).
- [2] N. Lee, R. T. Smith, A. Laine, *Interactive Segmentation for Geographic Atrophy in Retinal Fundus Images* IEEE Signal Processing Society, Asilomar Conference on Signals, Systems and Computers, (2008).
- [3] N. Lee, J. Caban, S. Ebadollahi, A. Laine, *Interactive Segmentation in Multi-Modal Medical Imagery using a Bayesian Transductive Learning Approach*, SPIE Medical Imaging Conference, To appear, (2009).
- [4] A. Bindewald, A.C. Bird, S.S. Dandekar et al., "Classification of Fundus Auto-fluorescence Patterns in Early Age-Related Macular Disease", *Invest. Ophthalmol Vis Sci.* 45, 3309-14, (2005).
- [5] Vapnik V. N., "Statistical learning theory", New York, Wiley, pp.339-371 (1998).
- [6] X. Zhu, "Semi-Supervised Learning Literature Survey", Computer Sciences Technical Report 1530, University of Wisconsin Madison, (2007).
- [7] S. Ebadollahi, J. Cooper, D. Kaufman, et al., "Concept-Oriented Access to Longitudinal Multimedia Medical Records: A Case Study in Brain Tumor Management", 3rd Int. Conf. on Semantics and Digital Media Technologies, 51-58, (2008).
- [8] R. Kindermann and J. L. Snell, "Markov Random Fields and Their Applications", (1980), American Mathematical Society.
- [9] S. Geman, D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 6, 721-741, (1984).
- [10] M. Jordan, "Learning in Graphical Models", MIT Press, (1998).
- [11] A. Ng, M. Jordan, "On Discriminative vs. Generative Classifiers: A comparison of logistic regression and Naïve Bayes", In NIPS 14 (2002).
- [12] K. Branson, "A Naïve Bayes Classifier Using Transductive Inference for Text Classification", Technical Report, Dept. of Computer Science and Engineering, UCSD (2001). Unpublished.
- [13] K. Nigam, A. McCallum, S. Thrun, and T. Mitchell, "Text Classification from Labeled and Unlabeled Documents using EM", *Machine Learning*, 39, 103-134, (2000).
- [14] D. Lowd, P. Domingos, "Naïve Bayes Models for Probability Estimation", *ICML*, 529-536, (2005).
- [15] A. Dempster, N. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *J. of the Royal Stat. Soc., Series B*, 39(1), 1-38, (1977).