LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-CONF-682239

AUTO-CONTEXT MODELING USING MULTIPLE KERNEL LEARNING

H. Song, J. J. Thiagarajan, K. N. Ramamurthy, A. Spanias

February 8, 2016

IEEE ICIP 2016
Phoenix, AZ, United States
September 25, 2016 through September 28, 2016

Disclaimer

AUTO-CONTEXT MODELING USING MULTIPLE KERNEL LEARNING

Huan Song[†], *Jayaraman J. Thiagarajan*[‡], *Karthikeyan Natesan Ramamurthy*[⋆], *and Andreas Spanias*[†]
[†] Arizona State University, Tempe, AZ
[‡] Lawrence Livermore National Labs, 7000 East Avenue, Livermore, CA
[⋆] IBM T.J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, NY

ABSTRACT

In complex visual recognition systems, feature fusion has become crucial to discriminate between a large number of classes. In particular, fusing high-level context information with image appearance models can be effective in object/scene recognition. To this end, we develop an auto-context modeling approach under the RKHS (Reproducing Kernel Hilbert Space) setting, wherein a series of supervised learners are used to approximate the context model. By posing the problem of fusing the context and appearance models using multiple kernel learning, we develop a computationally tractable solution to this challenging problem. Furthermore, we propose to use the marginal probabilities from a kernel SVM classifier to construct the auto-context kernel. In addition to providing better regularization to the learning problem, our approach leads to improved recognition performance in comparison to using only the image features.

Index Terms— Feature fusion, Marginalized kernel, Multiple kernel learning, Image classification

1. INTRODUCTION

In state of the art visual recognition systems, it is typical to adopt multiple descriptors (or features), which describe different aspects of the data. For example, in classical bag-of-words approaches, merging features from diverse cues such as shape, color and texture has been shown to improve the recognition performance. The success of feature fusion methods can be attributed to the use of a complementary set of features that can provide salient aspects for discriminating a large number of classes, while being robust to variations within a class. Though a variety of feature fusion technique exist in the computer vision literature, kernel methods provide a principled framework for fusing diverse descriptors into a unified feature space [1]. Commonly referred to as Multiple Kernel Learning (MKL) [2, 3, 4, 5, 6], this approach builds a Reproducing Kernel Hilbert Space (RKHS) for each descriptor and then fuses the multiple kernels as a non-negative linear combination [7]. or a hadamard product [8]. More recently, the use of randomized strategies for kernel construction has enabled the use of kernel methods in large scale [9], and in many applications they have been shown to perform as well as state of the art deep neural networks [10].

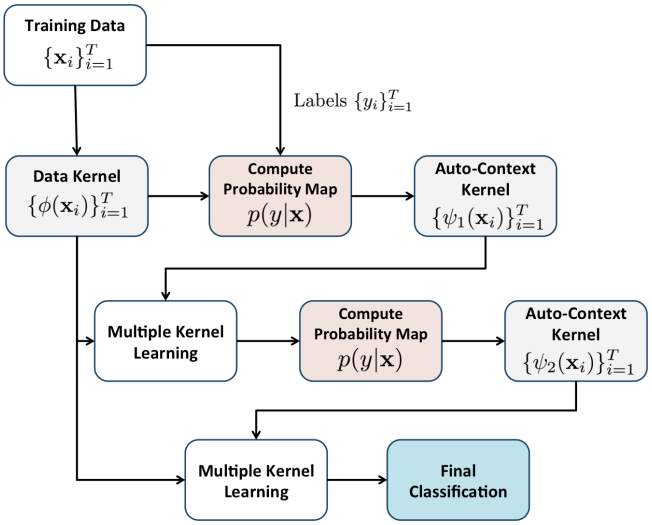

Fig. 1: Proposed approach for integrating auto-context with image features under the RKHS setting (illustrated for two iterations). This problem is solved efficiently by posing the fusion in each step as Multiple Kernel Learning (MKL).

A common characteristic of several existing feature fusion algorithms is that features employed are often low-level in nature i.e., they describe the local variabilities without taking the global context into account. However, it is known that high-level information, referred to as the context, is curcial to object/scene understanding. From a Bayesian viewpoint, context can be interpreted as the joint statistics of the multi-variate in the posterior probability, wherein the likelihood models the image appearance (observed data). In general, building a context model is challenging due to both the computational complexity in solving the MAP (Maximum A Posteriori) formulation and the difficulty in modeling complex patterns using limited training data. Consequently, *auto-context* models [11] have been developed, which can

This research was supported in part by the SenSIP center.

approximate the posterior using an iterative, supervisory approach. More specifically, these models integrate the low-level features with context information in the form of probability maps, obtained using a series of classifiers. By enabling the classifier to choose different supporting neighbors to modify the current probabilities towards the ground truth, auto-context methods lead to better regularization.

In this paper, we propose to adopt auto-context models under the RKHS setting. In addition to providing the flexibility of auto-context models, the proposed approach can build highly effective kernel models for object recognition. Since auto-context probability maps cannot be directly incorporated into the RKHS, we first estimate marginal probabilities using a classifier (e.g. Kernel Logistic Regression or Kernel SVM) and construct an *auto-context* kernel based on these probabilities. For example, the marginalized kernel construction in [12] can be used. Since any symmetric positive definite kernel defines a unique RKHS, we can use other forms of kernels by treating the probability map for each image as a feature vector directly. Interestingly, the process of fusing the auto-context model with the image appearance can be viewed as multiple kernel learning, for which a variety of efficient solutions exist. Figure 1 illustrates the proposed approach with two iterations. We demonstrate using standard object/scene classification datasets that the proposed approach results in highly effective recongition systems.

2. BRIEF REVIEW OF KERNEL METHODS

Let us consider the problem of binary classification using a Support Vector Machine (SVM) classifier that attempts to find a linear decision boundary between the two classes. When the classes are not linearly separable, it is beneficial to define a mapping function onto a high-dimensional space $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$ ($D > d$), such that SVM can yield a linear decision boundary in the resulting space. It is well known that this SVM formulation can be effeciently solved by considering its Lagrangian dual based on the kernel trick [1]. In other words, since finding the appropriate mapping function ϕ can be difficult, the dual formulation allows us to solve it solely based on the kernel matrices.

Definition Given the data domain $\mathcal{X} \subset \mathbb{R}^d$, a function $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is a valid kernel if it gives rise to a positive definite kernel matrix. i.e., $\mathbf{z}^T\mathbf{K}\mathbf{z} \geq 0, \forall \mathbf{z} \in \mathbb{R}^d$. In addition, a valid kernel defines an inner product and a lifting (transformation) ϕ, such that $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ where $\langle \cdot, \cdot \rangle$ denotes the inner product in the lifted space. This transformed space is referred as the reproducing kernel Hilbert space (RKHS).

Another interesting property of kernel methods is that fusing kernels from multiple sources (e.g. different features or representations) is straightforward. A commonly adopted strategy is to consider a convex combination of the kernels:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sum_m \beta_m k_m(\mathbf{x}_i, \mathbf{x}_j), \forall \beta_m \geq 0, \sum_m \beta_m = 1. \quad (1)$$

The process of simultaneously inferring the kernel weights $\{\beta_m\}$ and minimizing the structural risk (SVM objective) is referred to as Multiple Kernel Learning (MKL). For example, the SimpleMKL algorithm [13] solves a simplex constrained MKL formulation using its Lagrangian dual as follows:

$$\min_{\boldsymbol{\beta}} \max_{\boldsymbol{\alpha}} \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \sum_m \beta_m k_m(\mathbf{x}_i, \mathbf{x}_j)$$
$$\text{s.t. } \sum_i \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, \forall i, \sum_m \beta_m = 1, \boldsymbol{\beta} \succeq 0, \quad (2)$$

where α_i, α_j are the Lagrangian multipliers. As described earlier, solving this problem does not require the explicit knowledge of the mapping ϕ.

3. PROPOSED APPROACH

In this section, we describe the proposed approach for building an auto-context model in the RKHS, which is comprised of two main steps: (a) constructing the auto-context kernel, (b) integrating image features and the context model using multiple kernel learning.

Feature Extraction: Let us denote the dataset of N samples belonging to M different classes by $\{(\mathbf{x}^{(n)}, y^{(n)}), n = 1, ..., N\}$, where $\mathbf{x}^{(n)}$ and $y^{(n)} \in \{1, ..., M\}$ are the feature vector and the class label of the image n respectively. For simplicty, we adopt the popular bag-of-words model for building the feature representation. Assuming that there are S diverse descriptors, the visual word dictionaries of sizes $d_1, ..., d_S$ are learned using k-means clustering from the extracted descriptors. For a given image I_n, its feature representation is $\mathbf{x}^{(n)} = (x_1^{(n)}, ..., x_d^{(n)})$, where $d = \sum_s d_s$ is the feature dimension. Each feature component $x_j^{(n)}$ represents the normalized occurrence frequency of the j-th visual word (which is from the s-th dictionary) in the image n.

3.1. Constructing the Auto-Context Kernel

In order to construct the auto-context kernel, we begin by estimating the probability map for each image using a classifier in the RKHS. In particular, we propose to learn a kernel SVM, which can determine the relative importance of the visual words in classifying the image. In general, SVM classifiers predict only the class label without providing the probability information explicitly. Given M classes of data, for any sample $\mathbf{x}$, the goal is to estimate

$$p_i = P(y = i|\mathbf{x}), i \in 1, 2 \cdots M.$$

Adopting an one-vs-one classification scheme, we first estimate pairwise class probabilities r_{ij} by assuming that

$$r_{ij} = \frac{1}{1 + e^{A\hat{f}+B}},$$

where $\hat{f}$ is the decision value at $\mathbf{x}$. The parameters A and B are optimized by minimizing the negative log-likelihood of the training data. Upon estimation of the probabilities for all pairs of classes, we can consider the formulation in [14] to estimate the probabilities p_i.

$$\min_{\mathbf{p}} \frac{1}{2} \sum_{i=1}^{M} \sum_{j \neq i} (r_{ij} p_i - r_{ij} p_j)^2$$
$$\text{s.t. } p_i \geq 0, \sum_{i=1}^{M} p_i = 1. \quad (3)$$

This can be efficiently solved by considering its dual problem and using the iterative strategy proposed in [14]. Given the marginal probabilities, $p(y|\mathbf{x})$, for each image from SVM, we build the auto-context kernel as follows:

$$k_{AC}(\mathbf{x}_i, \mathbf{x}_j) = \psi(\mathbf{x}_i)^T \psi(\mathbf{x}_j),$$
$$= \sum_{y} \sum_{y'} p(y|\mathbf{x}_i, \boldsymbol{\gamma}_y) p(y'|\mathbf{x_j}, \boldsymbol{\gamma}_{y'}) S(y, y'), \quad (4)$$

where $y, y' \in \{0, 1\}$ and $S(y, y')$ denotes the similarity between the classes. Note, $k_{AC}(\mathbf{x}_i, \mathbf{x}_j)$ will result in a large similarity when the conditional probabilities that $\mathbf{x}_i$ and $\mathbf{x}_j$ belong to a class y is high. When the weighting term $S(y, y')$ is ignored, this corresponds to computing the linear kernel for the probability maps. Alternately, we can also construct a RBF kernel for the marginals.

3.2. Algorithm

The auto-context kernel measures how far the probability maps are from the ground truth. Consequently, this high-level information can effectively complement the image appearance information. Integrating the auto-context model into the feature kernel of the observed data is equivalent to fusing the two kernels, and we propose to solve this using multiple kernel learning. Denoting the RKHS corresponding to the image features and auto-context by $k_F(.,.)$ and $k_{AC}(.,.)$ respectively, the MKL formulation can be written as

$$k(\mathbf{x}_i, \mathbf{x}_j) = \beta_F k_F(\mathbf{x}_i, \mathbf{x}_j) + \beta_{AC} k_{AC}(\mathbf{x}_i, \mathbf{x}_j), \quad (5)$$

where $\beta_F, \beta_{AC} \geq 0$ and $\beta_F + \beta_{AC} = 1$. We use the SimpleMKL [13] algorithm to obtain the optimal coefficients. SimpleMKL performs optimization based on gradient descent on the SVM objective through a dual formulation. The overall iterative algorithm is described in Algorithm 1. The auto-context model can be progressively improved by learning a series of kernel classifiers using MKL. In each iteration, the marginal probabilities are estimated in the fused RKHS from the image feature kernel and the auto-context kernel from the previous iteration. Initially it is assumed that there is an uniform distribution, and hence auto-context has no useful information to improve the discrimination. Formally, the auto-context kernel for iteration t is constructed by learning a classifier using the fused kernel,

$$k^t(\mathbf{x}_i, \mathbf{x}_j) = \beta_F^{t-1} k_F(\mathbf{x}_i, \mathbf{x}_j) + \beta_{AC}^{t-1} k_{AC}^{t-1}(\mathbf{x}_i, \mathbf{x}_j). \quad (6)$$

Data: Image feature set $\{(\mathbf{x}^{(n)}, y^{(n)}), n = 1, ..., N\}$, where $y^{(n)} \in \{1, ..., M\}$, t_{max}
Result: Set of trained classifiers $\{H_t\}_{t=1}^{t_{max}}$
Build image feature kernel $\mathbf{K}_F$ using RBF;
Initialize $t = 1$, $\mathbf{K}^0 = \mathbf{K}_F$;
while $t \leq t_{max}$, *i.e., until preset number of iterations is not reached* **do**
  1. For each pairwise class, build a kernel SVM using the combined kernel $\mathbf{K}^{t-1}$ and store the classifier parameters in H_t;
  2. For each sample, estimate the probability map $\{p_i\}_{i=1}^{M}$ using (3);
  3. Construct the auto-context kernel $\mathbf{K}_{AC}^t$ based on the marginal probabilities using (4);
  4. Perform MKL to obtain the fused kernel $\mathbf{K}^t = \beta_F^t \mathbf{K}_F + \beta_{AC}^t \mathbf{K}_{AC}^t$ using (5);
  5. Set $t \rightarrow t + 1$;
end
Return the set of classifiers $\{H_t\}_{t=1}^{t_{max}}$;

Algorithm 1: Proposed algorithm for iterative estimation of auto-context in a RKHS setting.

4. EXPERIMENTS

In this section, we evaluate the proposed approach using standard visual recognition datasets and study the impact of auto-context modeling. The baseline comparison includes the case of using the feature kernel, and the auto-context kernel independently. Before we present the performance evaluation, we demonstrate the convergence behavior of the proposed algorithm in improving the marginal probabilities using the auto-context model.

4.1. Demonstration

The initial context model is equivalent to a uniform probability map with respect to all classes and hence the classification performance solely depends on the feature kernel. As the algorithm progresses, the auto context kernel will attempt to push the class probabilities closer to the ground truth using a series of classifiers. To illustrate this behavior, we consider a binary classification problem using a subset of the

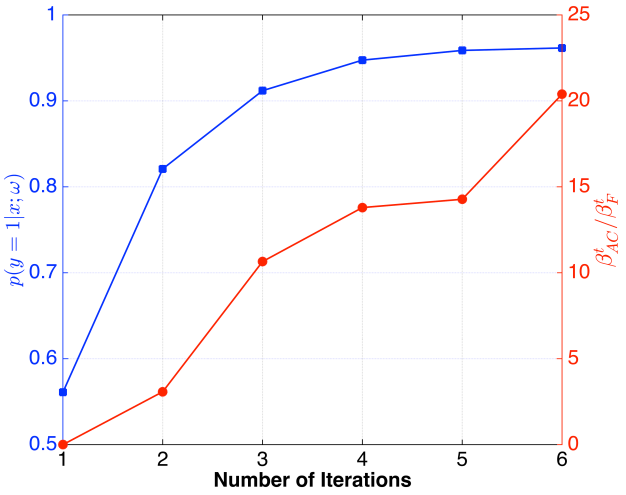

Fig. 2: Illustration of the convergence behavior of the proposed algorithm. Left axis shows the conditional probability of an example training sample, with ground truth $y = 1$, estimated by the kernel SVM classifier. Right axis shows the ratio of the importances between the auto-context kernel and the image feature kernel respectively.

Soccer dataset (details in the next section). Figure 2 illustrates how the relative importance of the auto-context kernel changes over the iterations, with respect to the image feature kernel. More specifically, we consider a training example with ground truth $y = 1$ and analyze the ratio of the weights, $\beta^t_{AC}/\beta^t F$. In addition, we plot the probability estimate from the kernel SVM, $p(y = 1|\mathbf{x})$. In the first iteration, the uniform context provides no additional information and hence the classifier is solely based on the feature kernel. However, as our algorithm proceeds, the auto-context kernel enables better discrimination between the two classes and hence $beta_{AC}$ becomes large. Interestingly, the marginal probability for the sample also changes from 0.55 in iteration 1 to 0.96 in iteration 6 indicating that the auto-context model leads to a more effective classifier.

4.2. Performance Evaluation

Soccer Dataset: This dataset contains images belonging to 7 soccer teams, comprised of 40 images per class. We used 25 images from each class for training and the 15 remaining images for testing. We extracted bag-of-words features based on both the shape and color cues. More specifically, the shape information was characterized by the SIFT descriptors computed at the set of keypoints determined by the Harris-Laplace point detector. The Hue-histogram [**?**], which decribed the color information, was evaluated at the same set of keypoints. Both descriptors are concatenated to construct the image appearance representation. The dictionary sizes for the SIFT and Hue descriptors were fixed at 400 and 300. The key points locations are determined by Harris-Lapalace key point detector. We constructed the RBF kernel for this image feature set with the parameter $\sigma = 50$. The kernel SVM classifier was designed with the parameter $C = 10$. Furthermore, the cost parameter and ℓ_2 regularization parameter for multiple kernel learning were set to 15 and 10 respectively. Table 1 shows the performance of our algorithm in comparison to baseline results obtained using only the feature kernel and the one-step auto-context kernel respectively. As the results indicate, the auto-context information enables the marginal probabilities to better match the ground truth in a few iterations, thereby leading to an improved recognition performance.

Table 1: Object recognition performance (% Accuracy) for standard datasets. We compare the performance of our algorithm against that obtained using only the image feature kernel and one step auto-context kernel.

Dataset	k_F+SVM	k_{AC}+SVM	Ours
Soccer	76.2	76.2	**81.9**
UCI Segmentation	86.9	85	**87.9**

UCI Image Segmentation Dataset: This dataset containts 2310 samples which were drawn randomly from a database of 7 outdoor images. The images were handsegmented to create a classification for every pixel. Each sample corresponds to a 3×3 regions. For the classification task, we extracted 19 different attribtues corresponding to the color statistics and used a RBF kernel with $\sigma = 10$. For multiple kernel learning, the ℓ_2 regularization was fixed at 0.1. As Table 1 indicates, incorporating the auto-context model improves the performance marginally. Note that, this dataset is comparatively easier to classify since the marginal probabilities of the training samples were close to the ground truth even after a single iteration.

5. CONCLUSIONS

In this paper, we presented a new approach for incoporating context modeling into a kernel learning formulation and showed that the fusion can be viewed as a multiple kernel learning formulation. By building a series of classifiers to approximate the target posterior, we demonstrated improvements in recognition perfomance. Future extensions to this work include designing randomized techniques for building auto-context kernels and exploring the use of other regularization strategies in feature fusion.

Prepared by LLNL under Contract DE-AC52-07NA27344

6. REFERENCES

[1] Bernhard Scholkopf and Alexander J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, MA, USA, 2001.

[2] Francis R Bach, Gert RG Lanckriet, and Michael I Jordan, "Multiple kernel learning, conic duality, and the smo algorithm," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 6.

[3] Mehmet Gönen and Ethem Alpaydın, "Multiple kernel learning algorithms," *The Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, 2011.

[4] Maria-Elena Nilsback and Andrew Zisserman, "Automated flower classification over a large number of classes," in *Computer Vision, Graphics & Image Processing, 2008. ICVGIP'08. Sixth Indian Conference on*. IEEE, 2008, pp. 722–729.

[5] Peter Gehler and Sebastian Nowozin, "On feature combination for multiclass object classification," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 221–228.

[6] Yi-Ren Yeh, Ting-Chu Lin, Yung-Yu Chung, and Yu-Chiang Frank Wang, "A novel multiple kernel learning framework for heterogeneous feature fusion and variable selection," *Multimedia, IEEE Transactions on*, vol. 14, no. 3, pp. 563–574, 2012.

[7] J.J. Thiagarajan, K.N. Ramamurthy, and A. Spanias, "Multiple kernel sparse representations for supervised and unsupervised learning," *Image Processing, IEEE Transactions on*, vol. 23, no. 7, pp. 2905–2915, July 2014.

[8] Jayaraman J. Thiagarajan, Karthikeyan Natesan Ramamurthy, Deepta Rajan, Andreas Spanias, Anup Puri, and David Frakes, "Kernel sparse models for automated tumor segmentation," *International Journal on Artificial Intelligence Tools*, vol. 23, no. 03, 2014.

[9] Ali Rahimi and Benjamin Recht, "Random features for large-scale kernel machines," in *Advances in Neural Information Processing Systems 20*, pp. 1177–1184. 2008.

[10] Po-Sen Huang, H. Avron, T.N. Sainath, V. Sindhwani, and B. Ramabhadran, "Kernel methods match deep neural networks on timit," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, 2014, pp. 205–209.

[11] Zhuowen Tu, "Auto-context and its application to high-level vision tasks," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[12] Koji Tsuda, Taishin Kin, and Kiyoshi Asai, "Marginalized kernels for biological sequences," *Bioinformatics*, vol. 18, no. suppl 1, pp. S268–S275, 2002.

[13] Alain Rakotomamonjy, Francis Bach, Stéphane Canu, and Yves Grandvalet, "Simplemkl," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.

[14] Ting fan Wu, Chih-Jen Lin, and Ruby C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2003.