A Context-Driven Forest-Based Approach to Hierarchical Scene Understanding

Zachary A. Daniels and Dimitris N. Metaxas

Motivation

- We wish to jointly solve several problems in scene understanding: *scene classification*, *scenario presence recognition*, *object presence recognition*, and *contextual priming*.
- Learning joint models for scene understanding is an active research problem, e.g. [4], [7], and [9].
- Existing methods typically use graphical models that operate on local image features, leading to good results at potentially high computational cost.
- We propose a model that is computationally efficient during training and test time.
- We introduce **Multi-Level Context Forests (MLCF)**, an extension of structured forests [1] to handle hierarchically-structured multi-label problems.
- Our MLCF model uses global image features (e.g. based on PlaceNet [10]) to make predictions about the content of scene images.
- We also introduce the concept of **scenarios**, sets of objects that commonly coexist, e.g. {toilet, shower, mirror, sink}.
- Scenes can be expressed as combinations of scenarios.
- Scenarios are flexible: objects can belong to multiple scenarios and a scenario can be present in a scene even if only a portion of its member objects are present.
- MLCFs exploit context by utilizing relationships within and between various levels of context.
- We examine a four-level contextual hierarchy: scenes → scenarios → objects → object organization/location.
- During training, we can use information about higher levels of context to restrict what information we need to examine when learning about lower levels of context.

References

[1] P. Dollar and C. Zitnick. Structured forests for fast edge detection. In Proceedings of the IEEE International Conference on Computer Vision, pages 1841–1848, 2013.
[2] H. Kim and H. Park. Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. Bioinformatics,23(12):1495–1502, 2007.
[3] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755):788–791, 1999.
[4] L.-J. Li, R. Socher, and L. Fei-Fei. Towards total scene understanding: Classification, annotation and segmentation in an automatic framework. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 2036–2043.IEEE, 2009.
[5] Y. Li and A. Ngom. The non-negative matrix factorization toolbox for biological data mining. Source code for biology and medicine, 8(1):1–15, 2013.
[6] R. Mottaghi, X. Chen, X. Liu, N.-G. Cho, S.-W. Lee, S. Fidler, R. Urtasun, and A. Yuille. The role of context for object detection and semantic segmentation in the wild. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2014.
[7] E. B. Sudderth, A. Torralba, W. T. Freeman, and A. S. Willsky. Learning hierarchical models of scenes, objects, and parts. In Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, volume 2, pages 1331–1338. IEEE, 2005.
[8] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In Computer vision and pattern recognition(CVPR), 2010 IEEE conference on, pages 3485–3492. IEEE, 2010.
[9] J. Yao, S. Fidler, and R. Urtasun. Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 702–709. IEEE, 2012.
[10] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning deep features for scene recognition using places database. In Advances in neural information processing systems, pages 487–495, 2014.

Scenarios

- Scenarios are sets of objects that commonly co-exist.
- Objects can belong to multiple scenarios and a scenario can exist in a scene instance without all of its member objects being present.
- Scenes can be expressed as combinations of scenarios.
- Scenarios are closely related to scene category.
- Scenarios can be learned from data using sparse non-negative matrix factorization [2][5].

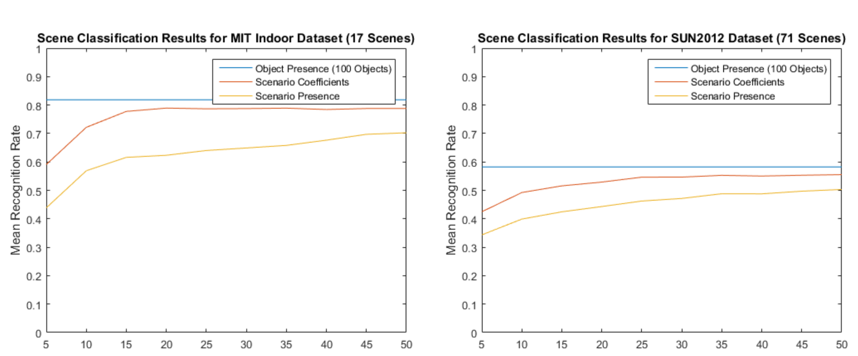
Examples of Scenarios

Scenarios for SUN2012 Dataset

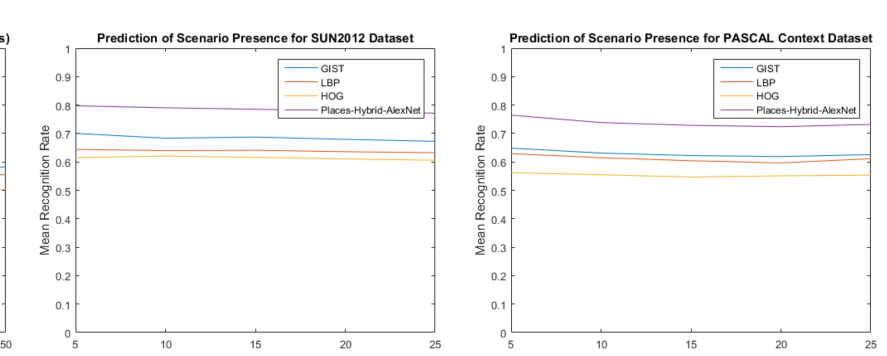
Grass	Bed	Desk	Chair	Worktop	Mountain	Faucet	Road	Sofa	Person
Tree	Desk Lamp	Screen	Table	Sink	Sea	Mirror	Car	Armchair	Ceiling
Plant	Night Table	Keyboard	Ceiling	Stove	Sky	Washbasin	Sidewalk	Coffee Table	Wall
Sky	Pillow	Mouse	Ceiling Lamp	Cabinet	Beach	Towel	Building	Cushion	Person Sit.
Ground	Curtain	Book	Floor	Oven	Rock	Countertop	Streetlight	Painting	Floor

Scenarios for PASCAL Context Dataset

Road	Monitor	Plate	Cat	Fence	Grass	Water	Train	Ceiling	Window
Car	Keyboard	Food	Dog	Horse	Ground	Boat	Track	Door	Sofa
Sidewalk	Mouse	Cup	Cloth	Saddle	Tree	Mountain	Platform	Floor	Curtain
Pole	Computer	Bottle	Bedclothes	Rope	Sky	Sky	Pole	Cabinet	Book
Bus	Paper	Table	Wall	Tree	Building	Bird	Ground	Chair	Potted Plant



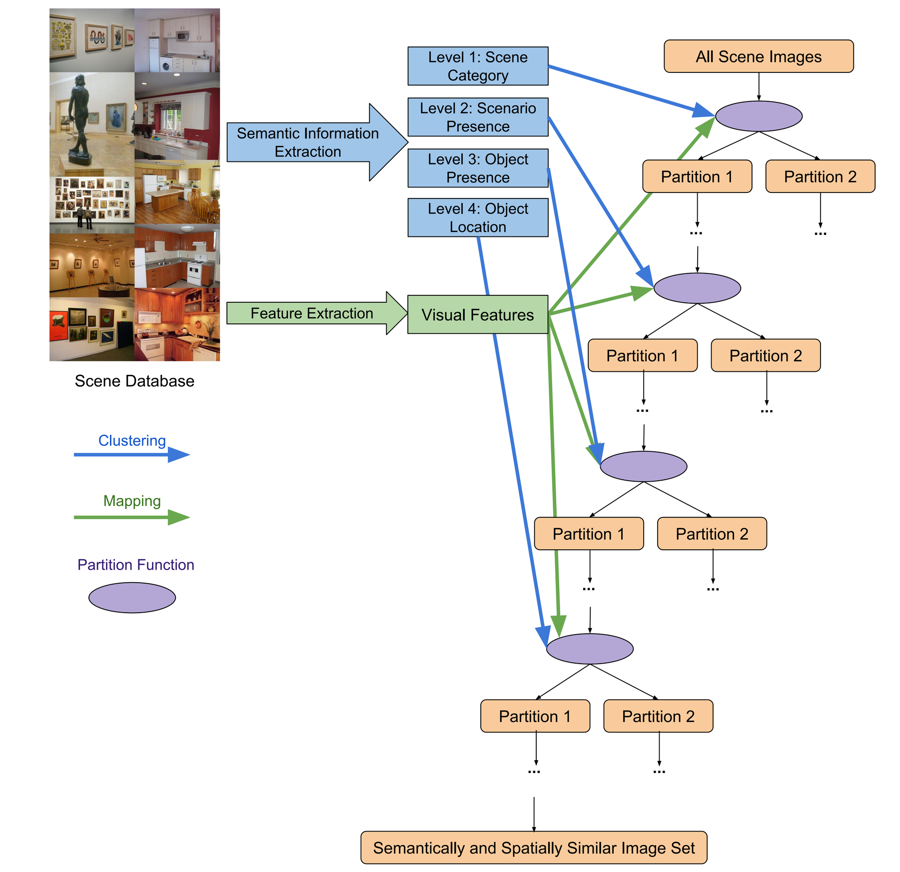
Scene Classification Using Scenarios



Scenario Recognition Using Visual Features

Multi-Level Context Forests (MLCF)

- MLCFs are constructed in a top-down manner, starting at the scene-category level and ending at the object-localization level.
- At each node in a MLC tree, we learn a splitting function by clustering a set of scene instances into two groups based on their labels and learning a separating hyperplane based on global image features.
- When the level of context switches, we identify the dominant "contextual objects" and restrict what labels are used when learning the splitting function for the subsequent level of context.
- Probabilities for every label in every level of context is stored at the leaf node.
- At test time, we extract global features and traverse the tree.



Example of MLCF for Retrieving Semantically and Spatially Similar Scene Images

Example of MLCF for Contextual Priming

Recognition: Macro F-Measure

	PASCAL Context		SUN2012			
Method	Scenarios	Objects	Scenarios	Objects	3 Scenes	16 Scenes
1-NN	0.464	0.425	0.494	0.476	0.501	0.126
5-NN	0.451	0.495	0.476	0.503	0.956	0.711
Linear SVM (Individual Classifiers)	0.391	0.471	0.390	0.441	0.957	0.605
Linear SVM (Individual Classifiers, Uniform Priors)	0.509	0.541	0.548	0.570	0.952	0.708
ML-5NN	0.474	0.520	0.505	0.526	0.958	0.743
ML-Naïve Bayes	0.208	0.395	0.155	0.285	0.904	0.000
BPMLL	0.412	0.524	0.439	0.528	0.947	0.589
MLCF	0.506	0.562	0.513	0.571	0.956	0.714

Localization

	PASCAL Context			SUN2012		
Method	IOU	Recall	Explored	IOU	Recall	Explored
1-NN: Semantic	0.331	0.390	0.131	0.307	0.336	0.113
5-NN: Semantic	0.386	0.538	0.202	0.380	0.519	0.176
1-NN: PlaceNet	0.188	0.240	0.108	0.190	0.238	0.108
5-NN: PlaceNet	0.295	0.543	0.274	0.305	0.565	0.256
Average: Semantic	0.315	0.851	0.501	0.308	0.883	0.457
MLCF: PlaceNet	0.330	0.798	0.375	0.333	0.845	0.398

RUTGERS

This work was supported by the National Science Foundation's Graduate Research Fellowships Program.