Sequential Bayesian Model Update under Structured Scene Prior for Semantic Road Scenes Labeling

Evgeny Levinkov, Mario Fritz
MPI Informatics

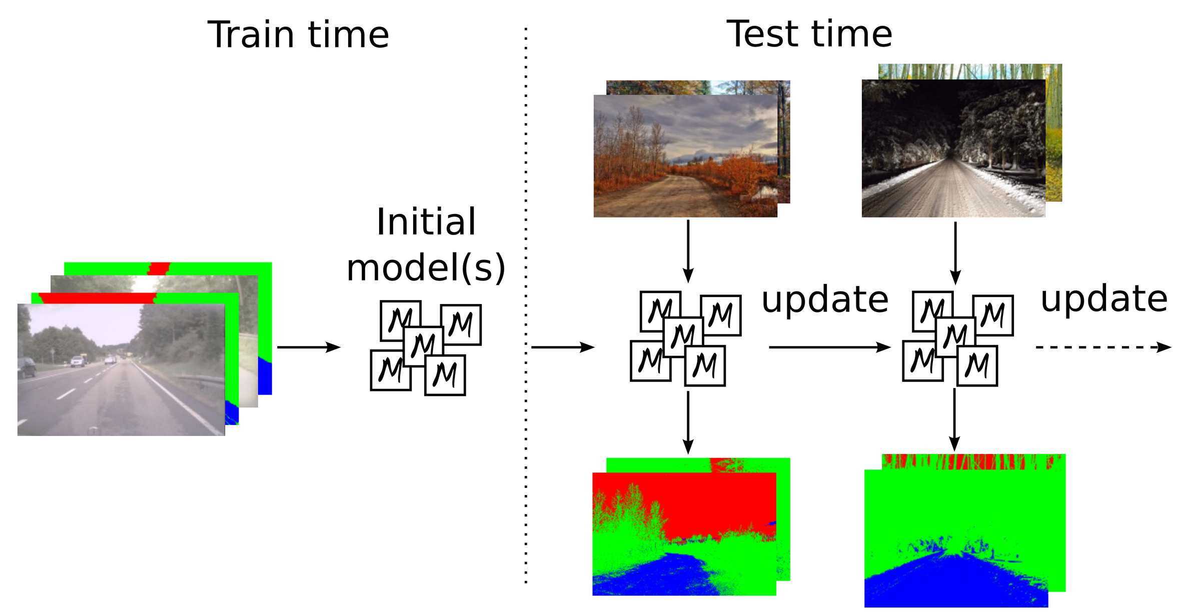
Motivation

We aim for vision systems that continuously operate in the real-world, where unforeseen conditions not represented in the training set are likely to occur. In order to equip systems with the ability to cope with such situations, we would like to enable adaptation to such new situations and conditions under certain assumptions.
Existing approaches:

- Non-adaptive approaches cannot account for changing feature distribution
- Domain adaptation techniques require at least some sample instances with ground truth labels from the target domain
- Global adaptive methods [1] cannot operate in continuous mode

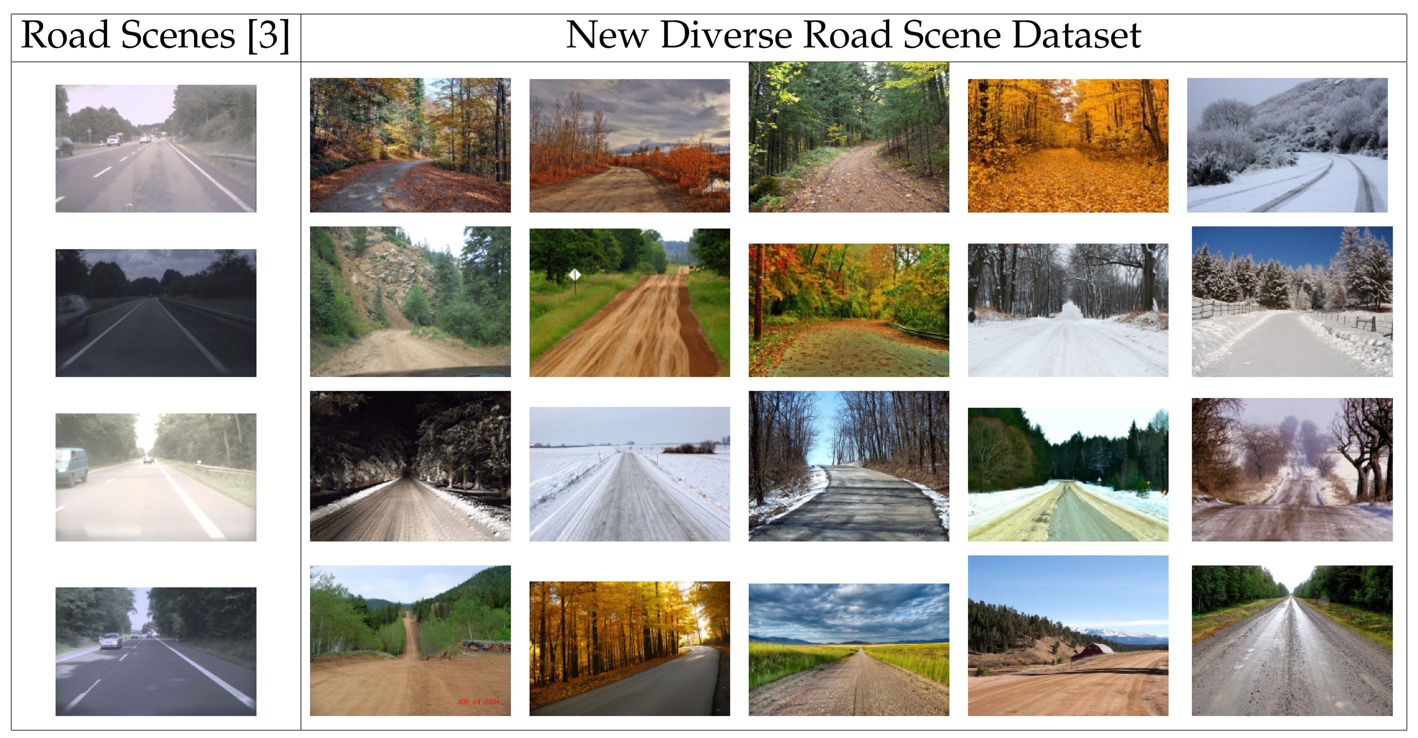
Approach

In order to robustly integrate new information at test time we propose a new Sequential Bayesian Model Update, which maintains a set of models via Particle Filter under the assumption of stationary label distribution



New Diverse Road Scenes Dataset

- We collected a new dataset of 220 images with road scenes representing wider range of visual conditions than before
- We used freely available images from Flickr®



- We use data set from [3] to train the initial model(s) on
- Comparison of Krähenbühl and Koltun [2] semantic image labeling algorithm on the old and the new test test

Test set	Fully connected CRF error, %			
	Road	Background	Sky	Average
Old	0.7	2.2	2.7	1.9
New	52.7	6.5	35	31.4

- Adaptation is necessary!
- Dataset and code for our method is available on-line

Naïve Model Update

Add new samples which confidence is higher than some threshold

- Label each lately arrived image using current model as $c^* = \operatorname{argmax}_{c \in \mathcal{Y}} P(x_{(i,j)} = c)$ for each pixel (i,j)
- Take features of only those pixels, for which $P(x_{(i,j)} = c^*) > \lambda$ holds, where λ is some predefined acceptance threshold parameter
- Retrain the model after a certain number of new images has been processed.

Naïve Model Update under Scene Prior

Add new samples which confidence is higher than some threshold w.r.t. the scene prior

- Compute histogram for each pixel on the training set and after per-pixel L_1-normalization get a prior $P_{pr}^{(i,j)}$ for each pixel $(i,j), i = 1, \ldots, W_{pr}, j = 1, \ldots, H_{pr}$
- At test time for an image with dimensions $W \times H$ compute $\tilde{P}(x_{(i,j)}) \propto P(x_{(i,j)}) P_{pr}^{(\lfloor i \frac{H_{pr}}{H} \rfloor, \lfloor j \frac{W_{pr}}{W} \rfloor)}$
- Take features of only those pixels, for which $\tilde{P}(x_{(i,j)} = c^*) > \lambda$ holds.

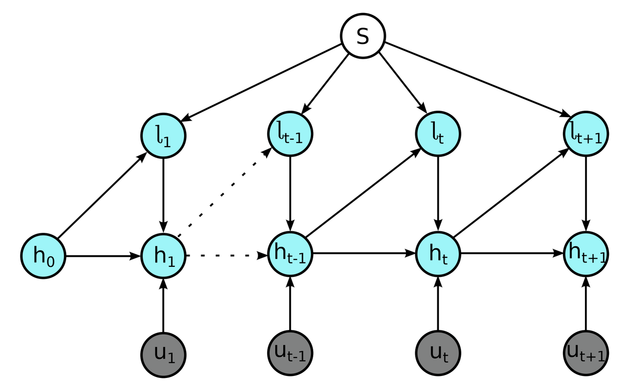
Sequential Bayesian Model Update under Structured Scene Prior

Maintain a set of particles with different characteristics, which evolve through time

- We maintain a population of models (particles $s_t^{(N)}$ with weights $\pi_t^{(N)}$) that approximate the distribution over the model-space $p(h_t|L_t)$
- Labeling X of a test image is done by marginalization over the model distribution

$$p(X|L_t) = \int p(X|h_t) p(h_t|L_t)\, \mathrm{d}h_t,$$

Bayesian Model Update - we are interested in modeling an evolving target distribution over models in order to account for the uncertainty in the unobserved scene labels. Therefore, we model the unobserved scene labels l_t of the unlabeled data u_t at time step t as a latent variable.



We describe the incorporation of the unlabeled examples in a Bayesian framework by integrating over all model hypothesis

$$p(h_t|L_{t-1}) = \int p(h_t|h_{t-1}, u_t) p(h_{t-1}|L_{t-1})\, \mathrm{d}h_{t-1}.$$

In the measurement step, we apply the Bayes' rule in order to get the updated distribution

$$p(h_t|L_t) = \frac{p(l_t|h_{t-1}, S) p(h_t|L_{t-1})}{p(l_t|L_{t-1})},$$

with

$$p(l_t|h_{t-1}, S) = p(l_t|h_{t-1}) p(l_t|S),$$

where $p(l_t|h_{t-1})$ is the probability of a certain scene labeling prediction given a model hypothesis h_{t-1} and $p(l_t|S)$ is a scene labeling prior.
Sampling - we propose to do model propagation by randomly choosing a subset of images which are provided to a particular classifier to retrain. For each particle i out of N:

1. Pick a particle s_t^i from $s_t^{(N)}$, which represents $p(h_t|L_t)$, according to the weights $\pi_t^{(N)}$
2. Sub-sample set of unlabeled images u_t to $\hat{u}_t$
3. Predict labels $\hat{l}_t = \operatorname{argmax}_l p(l|h_t)$ for subset $\hat{u}_t$
4. Accept or reject samples based on some threshold λ
5. Retrain model using $(\hat{u}_t, \hat{l}_t)$ and L_{t-1}

Directly normalize the weights of the particles $\pi_t^{(N)}$ to sum to 1.

Results

Setup and features:

- We performed training on the old set and testing on the new one
- For sequential adaptive algorithms we set the size of the batch to 10 images
- The order of images was randomly permuted and fixed for all tests
- We used features from [3] (Walsh-Hadamard transform, CIE-Lab color, .etc)
- As particles we used a Random Forest classifier consisting of 10 trees each having depth of at most 15 with 20% bagging

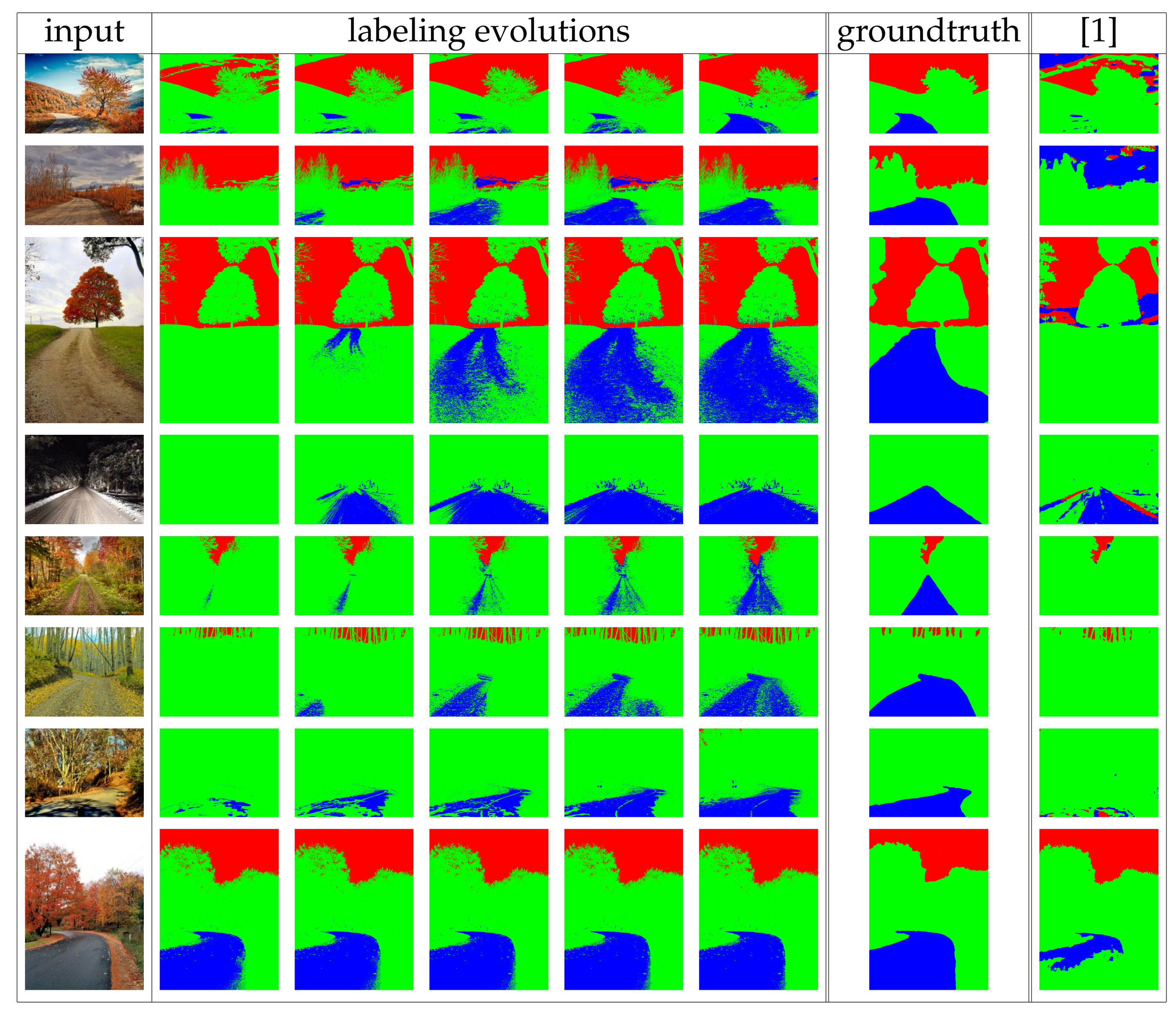
Numerical results of non-adaptive algorithms:

Method	Error, %			
	Road	Background	Sky	Average
Random Forest	43.8	**10.5**	29.1	27.7
Random Forest + FC-CRF [2]	40.8	11.0	**22.2**	**24.7**
Structured class-labels	**38.6**	11.1	33.1	27.6

Numerical results of adaptive algorithms:

Update type	Method	Error, %			
		Road	Background	Sky	Average
global	Alvarez *et al.* [1]	76.2	12.7	25.5	38.2
sequential	Naïve	26 ±1.4	**15.4±0.4**	9.3±1.4	17±0.7
	Naïve + Scene Prior	21±2.7	18.5±0.6	6.5±0.9	15.5±1.4
	Bayesian Model	**19±0.6**	18.3±0.6	**4.5±0.4**	**13.9±0.3**

Some visual results:



Acknowledgements

We would like to cordially thank Christian Wojek for advices and providing us with his code for feature extraction and José M. Álvarez for agreeing to run his method [1] on our new dataset.

References

[1] J. M. Álvarez, T. Gevers, Y. LeCun, and A. M. López. Road scene segmentation from a single image. In *ECCV*, 2012.
[2] P. Krähenbühl and V. Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *NIPS*, 2011.
[3] C. Wojek and B. Schiele. A dynamic conditional random field model for joint labeling of object and scene classes. In *ECCV*, 2008.