Bridging the Gap Between Synthetic and Real Data

Mario Fritz
Max Planck Institute for Informatics
Saarbrücken, Germany

June 25, 2015

There is a long tradition of using generative models in combination with discriminative classifiers [5, 6, 7]. Equally the recently successful deep learning technique [3] use jittering techniques [1, 2] that imply sampling from an underlying distribution. Although in both cases the the model is postulated and all parameters are in our control, we rarely achieve an accurate representation of the true underlying distribution. Yet, these techniques have shown improved performance as learning is guided by prior knowledge encoded in such generative models.

1 Learning and Prediction from Rendered and Synthesized Data

Many applications greatly benefit by means of synthesizing additional training data. For visual recognition this often involves a rendering process for creating new images. The employed model represents prior knowledge about the target domain. In this section, several examples are listed where we have directly used the rendered data – assuming that the domain mismatch between real and virtual examples is negligible.

Detection by Rendering. In early work, we have captured a light-field of an object and rendered new views of the object on demand in order to evaluate the posterior in a particle filter tracking framework [8].

New View Synthesis. Human generalize easily from a single view of an object to novel view-points. Today's computer vision algorithms are mostly learning and example based and therefore have to be shown variations across style and viewpoints in order to succeed. We have presented an approach that uses a 3D model to guide novel view synthesis, that is able to fill in disocculsion areas truthfully [9]. The object models trained on such augmented data show a greatly improved view point generalization.

Differentiable Vision Pipeline. Most recently, we have established a fully differentiable vision pipeline [10] that builds on top of an approximately dif-

ferentiable renderer [4] and a differentiated HOG image representation. This allows us to estimate object poses by exploiting the prescribed image synthesis procedure in the gradient computation.

2 Adaptation to Rendered and Synthesized Data

Although significant progress has been achieved by solely relying on realistic rendering and synthesis, quite often the domain shift between the virtual and the real world introduces a distribution mismatch that should be treated separately.
Visual Domain Adaptation via Metric Learning. We have proposed to reduce the effects of domain shifts by a metric learning formulation [11]. Hereby we have improved recognition across different data sources such a webcam, dslr or data from the web.
Recognition from Virtual Examples. We have employed the concept of metric learning for domain adaptation to the problem of visual material recognition [12]. The approach helps to bridge the gap between rendered and real data.
Prediction under changing prior distribution. Most recently, we have have shown how to perform gaze estimation in the wild [13]. Considering the change in the prior distribution of head pose and eye fixation distribution has been critical when training across datasets.

3 Unsupervised Adaptation

Future challenges include scenarios where no training data for adaptation is available. Less work has been performed in this direction. We have proposed to adapt to new conditions in a road segmentation task by assuming a stationary, structured prior over the label space, which allows us to successfully adapt a semantic labeler to unseen weather conditions [14]. Beyond the traditional recognition scenarios, we have also attempted to bring the required adaptivity to learning settings. E.g. we have adapted active learning strategies via reinforcement learning to different training distributions [15]. We hypothesize that non-parametric learning techniques for visual recognition and grouping [16] can be well suited to transfer structural relations across domains, while being less affected by changes in individual appearances.

References

[1] P. Simard, B. Victorri, Y. LeCun, J. Denker. Tangent prop-a formalism for specifying selected invariances in an adaptive network. In *Advances in neural information processing systems (NIPS)*, 1992

[2] D. Decoste, B. Schölkopf. Training invariant support vector machines. In *Journal of Machine Learning*, 2002

[3] A. Krizhevsky, I. Sutskever, G. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems (NIPS)*, 2012.

[4] M. Loper, M. Black. Opendr: An approximate differentiable renderer. In *European Conference on Computer Vision (ECCV)*, 2014

[5] T. Jaakkola, D. Haussler. Exploiting generative models in discriminative classifiers. In *Advances in neural information processing systems (NIPS)*, 1999

[6] M. Fritz, B. Leibe, B. Caputo, B. Schiele. Integrating representative and discriminant models for object category detection. In *IEEE International Conference on Computer Vision (ICCV)*, 2005

[7] A. Holub, M. Welling, P. Perona. Combining generative models and fisher kernels for object recognition. In *IEEE International Conference on Computer Vision (CVPR)*, 2005

[8] M. Zobel, M. Fritz, and I. Scholz. Object tracking and pose estimation using light-field object models. In *Vision, Modeling, and Visualization Conference (VMV)*, 2002.

[9] K. Rematas, T. Ritschel, M. Fritz, and T. Tuytelaars. Image-based synthesis and re-synthesis of viewpoints guided by 3d models. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[10] W.-C. Chiu and M. Fritz. See the difference: Direct pre-image reconstruction and pose estimation by differentiating hog. *arXiv:1505.00663 [cs.CV]*, 2015.

[11] K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting visual category models to new domains. In *European Conference on Computer Vision (ECCV)*, 2010.

[12] W. Li and M. Fritz. Recognizing materials from virtual examples. In *European Conference on Computer Vision (ECCV)*, 2012.

[13] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling. Appearance-based gaze estimation in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[14] E. Levinkov and M. Fritz. Sequential bayesian model update under structured scene prior for semantic road scenes labeling. In *IEEE International Conference on Computer Vision (ICCV)*, 2013.

[15] S. Ebert, M. Fritz, B. Schiele. Ralf: A reinforced active learning formulation for object class recognition In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[16] W.-C. Chiu, M. Fritz. Multi-class video co-segmentation with a generative multi-video model. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013