ZEISS

ZEISS SYMPOSIUM
OPTICS IN THE DIGITAL WORLD

WHITE PAPER ON COMPUTER VISION AND MACHINE LEARNING

Based on presentations and discussions during the respective workshop on 23 June 2016 in Oberkochen, Germany

---

Outreach of Computer Vision and Machine Learning

Computer Vision (CV) and Machine Learning (ML) have seen a tremendous evolution within the last 15 years. One of the main drivers of this success is the application of machine learning methods to computer vision tasks (image registration, segmentation, 3D reconstruction, tracking, object detection, image classification, …). These days it is widely agreed that difficult computational problems in data analytics (that cannot be solved analytically) are best solved with machine learning algorithms based on training data.

State-of-the-art

The current state-of-the-art CV allows the detection and tracking of single objects classes (such as faces, pedestrians or cars) in an unconstrained setting at a level that allows the realization of smart cameras that recognize smiling persons, driver assistance (pedestrian detection), surveillance applications and image-based web search [1, 2]. Image classification works on par with human level performance for databases with as many as 1000 classes (ImageNet) [3] and objects (e.g. birds) can be classified into fine-grained species with an accuracy of over 80% for 200 classes [4]. This level is sufficiently good for an app to support birders.

The field of structure-from-motion has reached performance levels that allow applications such as video editing and augmentation of large-scale 3D reconstruction from community web databases (e.g. Flickr) with the accuracy of a laser scanner [5]. The field of registration has reached maturity up to a level that allows photographs to be stitched seamlessly [6], e.g. from handheld cameras for panoramas.

Latest trends

During the last 5 years two lines of successful research have emerged: i) the integration of depth sensors (such as Microsoft Kinect, e.g. [7]) and ii) the application of deep learning techniques to basic computer vision tasks [8]. In particular the revival of deep learning methods improved the performance on many basic level tasks by leveraging large amounts of data in a learning framework. It has been agreed in the workshop that the next wave of innovation is likely to happen in the field of robotics where methods based on reinforcement learning can potentially model decision making processes.

On the computational side, the major trend is the advent of easily programmable interfaces for graphics processing units (GPUs). Interfaces such as CUDA or OpenCL are frequently used these days and allow the acceleration and parallelization of previously slow algorithms

up to frame-rate speed. In particular learning and evaluation of deep convolutional models is facilitated by GPUs. Undoubtedly the current success of deep learning methods would not have been possible without modern GPUs.

CV/ML in applications

Machine learning skills

The field of computer vision mostly evolves along the axes of robustness with respect to clutter and noise, runtime and performance. Common to most basic level tasks is the need for fast methods that parallelize well on modern hardware. Even though not trivial, this is mostly seen as an engineering issue for industry in the research community. However, as agreed in the workshop, the ability to engineer machine learning into product implementations is one of the key challenges for adaptation of machine learning methods in industry. The ability to engineer machine learning solutions comprises a diverse set of skills - ranging from a solid math background, modeling and optimization to efficient implementation skills (e.g. by leveraging GPUs). Many practicing software engineers, however, have a strong bias towards implementation skills.

Availability of data

Most current research is focused on 2D photographs that are easily accessible from the web. Consequently a lot of effort is spent on improving web applications. More research is required to evolve methods that are capable of processing multi-dimensional, multi-spectral and video data (e.g. for complex activity and event recognition, multi-modal registration). It has been agreed that this type of data certainly poses interesting and challenging scientific questions. A lack of (publicly) available data has been seen as the major showstopper for machine learning to be widely adapted in many domains. This is due to the fact that data is seen as an asset in many companies or research institutions and thus domain-specific data is often not released. Privacy and data safety is an additional issue that hinders data availability, in particular for medical application domains.

As we have seen substantial improvement by deep learning methods during the last 4 years and the necessity for large-scale, high-quality annotated data, there will be a continued high demand for a) large-scale datasets and b) large-scale learning and processing methods. Moreover, methods that allow existing knowledge to be reused from other domains might be an interesting direction of research (domain adaptation).

Algorithmic complexity

With respect to machine learning methods, learning algorithms need to be further improved. Current methods often require expert knowledge to achieve the best performance. For many applications, however, it will be necessary that users can train systems by themselves. Current automated training methods are often slow and non-automated methods require setting unintuitive hyper-parameters (e.g. deep learning methods). Thus there is a need for automated learning-methods with as few hyper-parameters as possible that require adjustment. A common assumption in the workshop was that in a few years machine

learning will be a commodity and even non-experts should be easily able to train machine learning systems.

Proposals

Datasets

The field is currently rapidly evolving and the commercial interest is ramping up substantially. However, at this point the field is mostly driven by IT companies (such as Google, Microsoft, Facebook, start-ups) and web or consumer applications. As the availability of large-scale, high quality datasets substantially influence the research community's direction (due to the widespread use of machine learning methods), there is a need for new datasets that define new applications (e.g. from the biomedical domains). In particular with respect to the development of smart products, established hardware vendors are in principle interested in evaluating computer vision and machine learning technology for their applications. Thus, hardware vendors together with their customers should collect and release (multi-sensor) data.

Challenges

Additionally they might provide challenges in the computer vision community in order to a) evaluate the current state-of-the-art for their domain and b) spur interest in further improving algorithms for their application. Successful initiatives in the computer vision community in this direction are the Middelburry benchmark (for the evaluation of stereo and optic flow algorithms) [10], the KITTI dataset (for evaluating technologies for autonomous driving) [9] and the ImageNet database (for the evaluation of image classification) [3].

Education

To enable a better mutual understanding of basic research and applications in industry, more joint projects will be required. One opportunity would be that more PhD projects or internships are offered to support academic research. This might in particular involve access to particular imaging devices (e.g. high-end microscopes). Another option is "industry on campus" initiatives that allow software engineers and scientists in industry to continuously learn on the latest algorithms in joint projects with academic institutions.

References

[1] M. Mathias, R. Benenson, M. Pedersoli, and L. Van Gool.
Face detection without bells and whistles.
European Conference on Computer Vision (ECCV), 2014.

[2] R. Benenson, M. Omran, J. Hosang, B. Schiele
ECCV workshop on computer vision for road scene understanding and autonomous driving

[3] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma,
Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei.
ImageNet Large Scale Visual Recognition Challenge.
arXiv:1409.0575, 2014

[4] S. Branson, G. Van Horn, S. Belongie, P. Perona
Bird Species Categorization Using Pose Normalized Deep Convolutional Nets
British Machine Vision Conference (BMVC), Nottingham, 2014.

[5] N. Snavely, SM Seitz, R. Szeliski
Modeling the world from internet photo collections
International Journal of Computer Vision, 80(2), pages 189-210, 2008

[6] M. Brown, D. Lowe.
Automatic Panoramic Image Stitching using Invariant Features.
International Journal of Computer Vision. 74(1), pages 59-73, 2007

[7] R. Newcombe, D. Fox, S. Seitz
DynamicFusion: Reconstruction and Tracking of Non-rigid Scenes in Real-Time,
Computer Vision and Pattern Recognition (CVPR), 2015

[8] Y. LeCun, Y. Bengio, G. E. Hinton
Deep Learning.
Nature, Vol. 521, pp 436-444

[9] A. Geiger, P. Lenz, R. Urtasun
Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite
Computer Vision and Pattern Recognition (CVPR), 2012

[10] D. Scharstein and R. Szeliski
A taxonomy and evaluation of dense two-frame stereo correspondence algorithms.
International Journal of Computer Vision, 47(1/2/3):7-42, April-June 2002.