EURASIP Journal on Image and Video Processing
a SpringerOpen Journal

EDITORIAL Open Access

Large-scale learning for media understanding

Anderson Rocha[1*†] and Walter J. Scheirer[2†]

1 Editorial

The remarkable growth in computational power over the last decade has enabled important advances in machine learning, allowing us to achieve impressive results across all areas of image and video processing. Numerical methods that were once thought to be intractable are now commonly deployed to solve problems as diverse as 3D modeling from 2D data (photo tourism [1]), object recognition (logo detection [2]), human biometrics (face recognition [3]), and video surveillance (automatic threat detection [4]). Despite this tremendous progress, there are many open questions related to understanding visual data—we are still far from matching human visual ability in all of these areas.

It is fair to scrutinize this observation in some detail—where are researchers and current methodologies falling short? From our perspective, the practicalities of real-world problems are often obscured by the mischaracterization of good results in limited contexts, theoretical frameworks built around artificial problems, and a deep sea of technical minutiae. Assumptions are a necessary component to problem solving, but poor ones lead our algorithms and subsequent analyses astray. Similarly, good theory is important, but we should not lose sight of the fundamental problem we are trying to solve by abstracting it away. Such circumstances occur more often than one might think.

How should a researcher charged with the design of an algorithm that must actually work outside of the laboratory avoid these dilemmas? Perhaps unsurprisingly, we, as academics, are typically more critical than constructive when evaluating new work—this is especially true of paper reviews and survey articles. In response to this, we submit the following researcher's guide to everyday machine learning as a gentle nudge forward:

1. **Do not make the problem easy.** The development of a capability for machines to understand scenes is a primary goal in computer vision. This problem is an exceedingly difficult one, requiring generalization to novel instances of known object classes amidst a practically infinite number of unknown object classes. However, to reach this goal, researchers have routinely targeted a much simpler problem; classification, which assumes that all object classes are known at training time. The difference in performance between an algorithm evaluated in this "closed set" regime and again in the actual "open set" one is dramatic. Recent work has shown that even when a data set as simple as the MNIST database of handwritten digits is re-contextualized into an open set problem where not all digits are known at training time, the performance of state-of-the-art supervised learning methods drops precipitously [5]. Other works have also shown the importance of open set classifier general recognition problems [6]. Therefore, always make sure to solve the original problem, and not one that is artificially easier.
2. **Test the perceptual thresholds of models.** A major shortcoming of current evaluation practices in visual learning is that they neglect an obvious frame of reference; that of the human observer. For example, the empirical gains achieved through deep learning architectures on benchmark data sets in computer vision, which have been characterized in the popular press as "sometimes even mimicking human levels of understanding" [7], are indeed impressive. However there is growing concern that such methods are actually inconsistent with human behavior, based on observed patterns of error [8]. Indeed, it is trivial to fool even the best deep learning algorithms into making mistakes humans never would by using a hill-climbing strategy that adds subtle distortions to an out-of-class image [9]. A better tactic is to follow the precise methods of visual psychophysics [10] and probe the perceptual thresholds of a model to understand its limits in a controlled manner. This

*Correspondence: anderson.rocha@ic.unicamp.br
†Equal contributors
[1] Institute of Computing, University of Campinas, Av. Albert Einstein, 1251, Cidade Universitária "Zeferino Vaz", 13084-971 Campinas, SP, Brazil
Full list of author information is available at the end of the article

will yield a quick answer as to whether or not it is consistent with human behavior.

3. **Move away from a strict adherence to data sets.** Related to the above observation, we have observed that posting good numbers on a benchmark data set is no longer a means to an end, but an end in itself. It is generally not true that a good result on a particular data set means that the algorithm which produced it will always perform well on images from outside of that data set. Well before all of the excitement over the performance of deep learning architectures on the ImageNet challenge [11], Torralba and Efros [12] questioned the field's singular focus on such narrow problems, arguing that all data sets in computer vision contain some measure of easily learned bias that can inevitably lead to false conclusions. Bias becomes evident when testing an algorithm's cross-data set generalization ability, or, in other words, training a model on one data set and applying it to another. We recommend that researchers go even further by testing their algorithms on data from sources external to any data set—if an algorithm fails when presented with frames from a live camera, more work needs to be done.
4. **Avoid dogma (but do so in a principled way).** Like any academic field, machine learning has its share of subdisciplines, each with its own prescriptions for problem solving. Sometimes, these subdiscipline-specific views become stumbling blocks to general progress. An example of this is the topic of convex optimization, which has come to be the dominant mode of optimization for visual recognition problems. More often than not, it is frowned upon to propose an algorithm that may get trapped in local minima—even if it demonstrates superior empirical performance over the state-of-the-art. Thankfully, the reemergence of artificial neural networks, which are non-convex, has loosened this tension by demonstrating the utility of complex and hierarchical network structures that are not amenable to convex optimization [13]. Hence, strive to design an algorithm that works to your performance specification, and not one that is unnecessarily constrained by theory. However, if a theory does lead you to a good solution for particular cases, take advantage of it.
5. **Seek different evidence when characterizing visual data.** There is no silver bullet to solve all problems—especially when describing images. Different problems often demand different forms of image description. However, even within a single problem, it is hard to think of a simple descriptor that captures all the nuances and cues present in an image. Consider the example of content-based image retrieval. Using a color descriptor is not enough to capture all possible class variability. Including other complementary features, such as shape and texture, is key for a successful retrieval system. In a remote-sensing image-classification system, the RGB color channels are just one way to capture image information. Infrared channels can also play an important role, and each channel can have its own custom-tailored descriptors. Therefore, we recommend thinking of possible complementary features when dealing with visual problems, along with innovative ways for combining them. Sometimes what seems unsolvable using just one piece of visual evidence becomes much easier when considering evidence from different and complementary features and sensors.
6. **Be aware of machine-learning black-boxes.** With the ever-increasing need for processing vast amounts of data, researchers often rely on off-the-shelf machine learning solutions to tackle their problems using so-called black-boxes. Although it is quick and easy to turn to such solutions, this comes at a price; if the underlying problem is poorly understood, the default parameters of a chosen black-box model will likely result in poor performance. Hence, we recommend that researchers pay close attention to the intrinsic properties of their problems and to carefully choose the learning algorithm and its parameters when actually implementing a solution. Sometimes just a small amount of parameter tuning can save weeks of processing and yield very good classification results.
7. **Think of new useful applications.** Researchers these days concentrate on just a handful of well-known applications. Digital photo tagging is, without a doubt, a great application, but it is not the only one we should be working on. Get creative when demonstrating the capabilities of a new algorithm. Some interesting applications that we have seen lately include the following: shellfish detection for the protection of fisheries [14], digital restoration of historical documents [15], and steering headlight beams around raindrops [16]. These are a good start, but there is certainly much more over the horizon.

With the above advice setting the stage, this special issue examines emerging questions and algorithms related to complex visual processing tasks where machine learning is applicable. This spans a number of important problems at multiple stages of the image analysis pipeline, from features to decision-making strategies, all the way through to end-user applications. This issue brings together seven articles describing original research that is closely matched to these stages.

At the cusp of current visual recognition capabilities are algorithms that learn features, instead of just blindly applying hand-tuned features that are not domain specific. In "Hyperspectral Image Classification via Contextual Deep Learning," Ma et al. examine the applicability of this paradigm to hyperspectral imaging, and report promising results for classification in remote sensing, where this modality is commonly deployed. This article presents a highly effective, yet highly parameterized learning-based framework—what options do we have to tune it? In their article "On the Optical Flow Model Selection Through Metaheuristics," Pereira et al. propose the use of methods from the area of evolutionary computing to optimize parameter sets during training to minimize error. The results after searching the parameter space this way are remarkably better.

Making a good decision is just as critical as learning a good representation. Chen et al. introduce us to a new scalable strategy for large-scale learning in "A Robust SVM Classification Framework Using PSM for Multi-class Recognition." The final supervised classification step of an image analysis pipeline is often a bottleneck—robustness and speed in the overarching framework are the key to solving this, according to Chen et al.

As any practitioner of machine learning knows, there are some situations in which we do not have labels for all of our training data. A viable solution in such a case is to learn from labeled and unlabeled images via semi-supervised learning. In "A Semi-Supervised Learning Algorithm for Relevance Feedback and Collaborative Image Retrieval," Pedronette et al. highlight the power of this approach for communities of users participating in collaborative image retrieval.

Once we have tools that can learn over large amounts of data, a host of previously unapproachable problems can be solved. There is, for instance, an immediate need for medical imaging algorithms that can support doctors in making accurate diagnoses. In "Oriented Relative Fuzzy Connectedness: Theory, Algorithms, and its Applications in Hybrid Image Segmentation Methods," Bejar and Miranda describe a new method for segmentation that is applied to brain and chest images from MRI and CT scans. And, at the cellular level, Xu et al. target the identification of antinuclear antibodies as evidence of autoimmune diseases with their work "HEp-2 Cells Classification Based on a Linear Local Distance Coding Framework". And finally, Islam et al. take us on a global tour to witness the regional diversity and often surprising cultural homogeneity of the human face, as represented by models built from geo-tagged face images in their article "Large-Scale Geo-Facial Image Analysis."

We hope that you enjoy this special issue as much as we did while putting it together.

Acknowledgements
The editors would like to thank all of the reviewers for their hard work and insightful comments, which have allowed us to assemble an outstanding special issue. Their contribution is much appreciated. Prof. Anderson Rocha also thanks the financial support of the Brazilian Coordination for the Improvement of Higher Level Education Personnel (CAPES) through the DeepEyes project.

Author details
[1]Institute of Computing, University of Campinas, Av. Albert Einstein, 1251, Cidade Universitária "Zeferino Vaz", 13084-971 Campinas, SP, Brazil. [2]Department of Computer Science and Engineering, University of Notre Dame, Fitzpatrick Hall of Engineering, 46556 Notre Dame, Indiana, USA.

Received: 14 July 2015 Accepted: 15 July 2015
Published online: 30 July 2015

References

1. K Matzen, N Snavely, in *ECCV*. Scene chronology, (2014)
2. R Pandey, D Wei, V Jagadeesh, R Piramuthu, A Bhardwaj, in *IEEE ICIP*. Cascaded sparse color-localized matching for logo retrieval, (2014)
3. Y Taigman, M Yang, M Ranzato, L Wolf, in *IEEE CVPR*. Deepface: closing the gap to human-level performance in face verification, (2014)
4. P Turaga, R Chellappa, VS Subrahmanian, O Udrea, Machine recognition of human activities: a survey. IEEE T-CSVT. **18**(11), 1473–1488 (2008)
5. LP Jain, WJ Scheirer, TE Boult, in *ECCV*. Multi-class open set recognition using probability of inclusion, (2014)
6. WJ Scheirer, A Rocha, A Sapkota, TE Boult, Toward open set recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). **35**(7), 1757–1772 (2013)
7. J Markoff, *Researchers announce advance in image-recognition software*. (The New York Times, 2014)
8. C Szegedy, W Zaremba, I Sutskever, J Bruna, D Erhan, I Goodfellow, R Fergus, in *International Conference on Learning Representations (ICLR)*. Intriguing properties of neural networks, (2014)
9. C-Y Tsai, DD Cox, Are deep learning algorithms easily hackable? http://deeplearning.twbbs.org/. Accessed March 16 2015
10. Z-L Lu, B Dosher, *Visual psychophysics: from laboratory to theory*. (The MIT Press, Cambridge, MA, 2013)
11. O Russakovsky, J Deng, H Su, J Krause, S Satheesh, S Ma, Z Huang, A Karpathy, A Khosla, MS Bernstein, AC Berg, L Fei-Fei, Imagenet large scale visual recognition challenge. CoRR (2014). abs/1409.0575
12. A Torralba, AA Efros, in *IEEE CVPR*. Unbiased look at dataset bias, (2011)
13. Y Bengio, Y LeCun, in *Large Scale Kernel Machines*. Scaling learning algorithms towards AI (MIT Press Cambridge, MA, 2007), pp. 321–358
14. M Dawkins, C Stewart, S Gallager, A York, in *IEEE WACV*. Automatic scallop detection in benthic environments, (2013)
15. K Pal, C Schüller, D Panozzo, O Sorkine-Hornung, T Weyrich, Content-aware surface parameterization for interactive restoration of historical documents. in Computer Graphics Forum (Proc. Eurographics). **33**(2) (2014)
16. R Tamburo, E Nurvitadhi, A Chugh, M Chen, A Rowe, T Kanade, SG Narasimhan, in *ECCV*. Programmable automotive headlights, (2014)