SPECIALTY GRAND CHALLENGE
published: 18 July 2022
doi: 10.3389/fimag.2022.951926

Grand Challenges in Imaging Applications

*Alessandro Piva**

Department of Information Engineering, University of Florence, Florence, Italy

Keywords: image processing, video processing, imaging applications, coding standards, robustness, reproducibility, explainability, datasets

Edited and reviewed by:
Jinchang Ren,
Robert Gordon University,
United Kingdom

****Correspondence:***
Alessandro Piva
alessandro.piva@unifi.it

Specialty section:
This article was submitted to
Imaging Applications,
a section of the journal
Frontiers in Imaging

***Received:** 24 May 2022*
***Accepted:** 23 June 2022*
***Published:** 18 July 2022*

Citation:
Piva A (2022) Grand Challenges in
Imaging Applications.
Front. Imaging. 1:951926.
doi: 10.3389/fimag.2022.951926

1. INTRODUCTION

It is my honor to be the inaugural Specialty Chief Editor of Imaging Applications, one of the sections of the new journal Frontiers in Image Processing. This section features original research work, tutorial and review articles and welcomes submissions from academic and industry researchers that seek to develop new applications of image and video processing. Topics of interest covered by this section include, but are not limited to:

- Security (Data hiding, steganography, forensics, encryption)
- Biometrics
- Surveillance
- Cultural Heritage
- Remote sensing
- Document processing
- Astronomy
- Automated driving
- Robotics
- Gaming
- Sports

The general field of image and video processing has witnessed in the last few years the flowering of ever new applications that have affected countless areas of our life, thanks particularly to advancements in computational performance, transmission bandwidth availability, design of new acquisition devices and development of effective Artificial Intelligence (AI) tools (Dufaux, 2021). In this paper, we highlight a few research grand challenges for imaging applications, in order to attract the attention of researchers to this research area. Given the hugeness of the field, this list is by no means exhaustive.

2. ADOPTION OF NEW IMAGE AND VIDEO CODING STANDARDS

The most widely adopted video coding standard is currently the H.264/MPEG-4 Advanced video coding (AVC) (Wiegand et al., 2003), which was initially developed between 1999 and 2003, and was then extended in 2003–2009. So far, most hardware manufacturers support it, such that it is still a fundamental video technology in a wide range of video applications, including almost all the major platforms for video streaming. After AVC, the other adopted standard is the High Efficiency Video Coding (HEVC) standard (Sullivan et al., 2012), that was finalized in

2013, providing about a 50% bit-rate reduction compared with AVC standard. The efficiency of current video coding standards, like H.264 and HEVC, is still not sufficiently high for today's heterogeneous data-intensive multimedia applications, that often work in a wireless transmission environment or with limited computational resources. Thus, there is an ever-increasing request for developing new video coding standards showing a high compression ratio and, at the same time, high visual quality (Zhang and Mao, 2019). This request can be alleviated by the fact that the available computational power keeps rising, thanks to the adoption of parallel computing together with hardware acceleration, thus allowing to adopt more complex and effective coding algorithms, even if still based on the principles of transform coding and predictive coding. At this aim, the Versatile Video Coding (VVC) standard has recently been finalized (Bross et al., 2021), achieving approximately another 50% bit rate reduction for the same subjective quality when compared to HEVC. However, the adoption of new standards is very limited in most imaging applications; there are several motivations, one reason is that these standards are ruled by several patent pools with different pricing structures and terms and conditions; another reason is the fact that very often, in real application scenarios the available computing power is not sufficient for an effective decoding process; in addition, in many cases there are concerns about backward compatibility with previous versions of the codecs. So, there is still plenty of space for the adoption of those recent standards, that remain confined to very limited applications. A similar situation holds for the case of image coding standars: here, JPEG (Hudson et al., 2017) is currently the most widely used lossy image compression standard, although it was introduced in 1992. Its successors, starting from JPEG 2000 (Taubman and Marcellin, 2002), to the High Efficiency Image File Format (HEIF) (Lainema et al., 2016), and to the AV1 Image File Format (AVIF), which is the latest image compression standard, have not been sufficiently spread in imaging applications.

3. ROBUSTNESS OF DEEP LEARNING-BASED SOLUTIONS

Imaging applications have all seen the technological migration from model-based algorithms to data driven-based algorithms. A model-based approach, relying on some mathematical or statistical models of the application scenario and the involved data, has usually good performance, provided that the application follows the designed model, whereas it has an in most cases a smooth degradation of performance if this model does not correctly represent the real world scenario. More recent methods, instead, are for the most part data-driven: they exploit the availability of large amounts of data to learn how to solve the problem at hand. These methods, relying on Convolutional Neural Networks (CNN) and other Deep Learning (DL) architectures (LeCun et al., 2015), outperform at large the previous methods, such that they have been adopted in almost all applications requiring imaging technologies. However, these methods still present several limits that need to be addressed carefully in order to be applied in real life imaging applications.

Deep learning has the necessity of training the tools on a huge amount of data which is representative of the variety of situations encountered in real-life applications, resulting in a strong training inefficiency (Strubell et al., 2019); this is due to the fact that performance depends heavily on the alignment between training set and test data: it is extremely good when training and test sets are matched, but if unrelated datasets are used at test time, it usually drops to values near to random guess; indeed, generalization, that is the ability to perform well on unseen data, is one of the key elements of machine learning (Bishop and Nasrabadi, 2006). The necessity of dealing with situations that could not be foreseen at training time is also a typical problem in many applications. The risk that deep learning based tools are overfitted to the training data and fail to give a correct answer in the presence of new unforeseen situations must be carefully considered. This lack of robustness thus can limit the applicability of learning based approaches to specific scenarios. In addition, not only the dataset has to be enough big, but data need also to be fully labeled, at least for supervised learning, the most common form of deep learning. A new solution is represented by self-supervised learning (Goyal et al., 2019), that is algorithms that are able to learn from large-scale unlabeled data, without the need for manual annotations, but still the research in this area is not spread over many scenarios. It is then evident that the process of creating a training dataset properly adapted to a particular imaging application could be really cumbersome, as will be detailed in another section of this work.

In several imaging applications, with particular focus on security-related fields, it has to be taken into account the possible presence of an adversary which could actively try to mislead the analyses. In fact, a skilled user, aware of the principles on which deep learning tools rely, may purposely apply some measures to force wrong output of the methods. Indeed, the research has shown that it is rather easy to generate the so-called adversarial examples, that is inputs obtained by applying small but intentionally worst-case perturbations to examples from the dataset, such that the perturbed input results in the model outputting an incorrect answer with high confidence (Goodfellow et al., 2014; Papernot et al., 2016). Even worse, some studies demonstrated the transferability of such attacks to algorithms different than those directly targeted by the attack (Liu et al., 2016). A competition between attacks and defenses for adversarial examples is occurring in the research field (Yuan et al., 2019), such that understanding and possibly ensuring the resilience of deep learning tools against attacks is a crucial problem, if they have to be used under the intrinsically adversarial conditions typical of several imaging applications.

4. REPRODUCIBILITY AND EXPLAINABILITY OF ALGORITHMS

The reproducibility of research, intended as the possibility to accurately reproduce the results of an experiment described in a paper, is a key principle of computing science (Schwab et al., 2000). It has been demonstrated that in the field of signal processing—and thus of imaging applications also—the reproducibility can be achieved not by just providing a detailed description of the proposed algorithm in a published work, but by also making available the source code of the algorithm, and the measurement data, along with the details of all parameter settings (Vandewalle et al., 2009). Even if the last years an improved effort has been documented in this direction (Vandewalle, 2019), still a lot of work will be required in most imaging application, by taking into account that the diffusion of deep learning-based models makes almost impossible the reproducibility of any experiment without the sharing of the complete source code along with the related training and test datasets.

The black-box nature of deep learning methods makes it difficult to interpret the results of the analysis and understand the motivations of a particular decision made by the algorithm, and how it relates to particular characteristics of the input data (Zhang and Zhu, 2018). These characteristics represent a strong limitation in many decision-critical applications: expert users, that is the software designers and developers, will require explanation tools that allow them to understand the behavior of the implemented methods for different setting conditions; end users will need access to a satisfactory explanation for the process that led to the decision: if these conditions are not met, a sufficient trust on the automatic system will not be easily achieved, hindering its adoption in real case scenarios (Ras et al., 2018). Indeed, in the deep learning area there are several emerging trends supporting explainability of decisions, like visualization methods, model distillation and intrinsic methods (Ras et al., 2022). However, explainable deep learning is still in its early phase and more developments are needed, since current solutions of explaining deep learning are seldom enough to achieve explanations helpful in practice, still impeding to deploy and liably exploit deep learning-based solutions, for instance in application domains such as autonomous driving or facial recognition. At this aim, it has to be taken account that according to the particular imaging application, a different purpose and type of explanation will be needed, so that it is not obvious what the best type of explanation metric should be adopted for each particular scenario (Gilpin et al., 2018).

5. AVAILABILITY OF PUBLIC DATASETS

Another fundamental aspect for the development of next imaging applications is the availability of large and well designed public datasets. In the last years, there has been the deployment of large datasets in the computer vision area; the most adopted one is probably the ImageNet (Deng et al., 2009), containing over 15 millions labeled images divided into over 22,000 classes. Starting from it, a subset defined ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015) has been derived, spanning 1000 object classes and containing 1,281,167 training images, 50,000 validation images and 100,000 test images. These very large datasets are extremely effective for training general purpose networks, but present some issues like the question of consent and privacy (Birhane and Prabhu, 2021). Notwithstanding the presence of these general datasets, new emerging imaging modalities or specific application scenarios require the deployment of their specific datasets, but this process is hindered by the difficulties in collecting a sufficient number of relevant data, and by the fact that the underlying imaging technologies are continuously evolving. Let us provide an example of this problem, related to the multimedia forensics area. In that field, one of the research issues is the identification of the acquisition device that generated the content, image or video; one of the most adopted features is the Photo Response Non-Uniformity (PRNU), a unique pattern noise left by each sensor (Lukas et al., 2006; Iuliani et al., 2019). Its uniqueness ensures that the sensor pattern noises extracted from different cameras are strongly uncorrelated, even when they belong to the same camera model. To test the performance of the methods based on this feature, several datasets have been proposed, the most adopted one being the VISION dataset (Shullani et al., 2017). The VISION dataset is currently composed by 34,427 images and 1,914 videos, both in the native format and in their social version (Facebook, YouTube, and WhatsApp are considered), from 35 portable devices of 11 major brands. The limitation of this dataset is that these devices have been released several years ago, so their imaging technology is some- what obsolete. In the meantime, with the advent of computational photography, the image acquisition pipeline is rapidly changing with novelties involving both the acquisition process, and the in-camera processing, and this process could hinder the usefulness of PRNU. Preliminary analysis carried out on over 33.000 Flickr images belonging to 45 smartphone and 25 DSLR camera models released recently, show that non-unique artifacts that may reduce PRNU noise's distinctiveness, especially when several exemplars of the same device model are involved in the analysis, appear in some of the most recent devices (Iuliani et al., 2021). These results raise awareness about a possible issue with PRNU reliability, especially in the law enforcement world, and thus it is essential to keep validating such technology on recent devices as they appear, requiring the creation of updated datasets.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Birhane, A., and Prabhu, V. U. (2021). "Large image datasets: a pyrrhic win for computer vision?" in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)* (Waikoloa, HI: IEEE), 1536–1546.

Bishop, C. M., and Nasrabadi, N. M. (2006). *Pattern Recognition and Machine Learning, Vol. 4*. New York, NY: Springer.

Bross, B., Chen, J., Ohm, J.-R., Sullivan, G. J., and Wang, Y.-K. (2021). Developments in international video coding standardization after avc, with an overview of versatile video coding (vvc). *Proc. IEEE* 109, 1463–1493. doi: 10.1109/JPROC.2020.3043399

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). "Imagenet: a large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (Miami, FL: IEEE), 248–255.

Dufaux, F. (2021). Grand challenges in image processing. *Front. Signal Process*. 1, 664232. doi: 10.3389/frsip.2021.664232

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., and Kagal, L. (2018). "Explaining explanations: An overview of interpretability of machine learning," in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)* (Turin: IEEE), 80–89.

Goodfellow, I. J., Shlens, J., and Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*. doi: 10.48550/arXiv.1412.6572

Goyal, P., Mahajan, D., Gupta, A., and Misra, I. (2019). "Scaling and benchmarking self-supervised visual representation learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (Seoul: IEEE), 6391–6400.

Hudson, G., Léger, A., Niss, B., and Sebestyén, I. (2017). Jpeg at 25: Still going strong. *IEEE Multimedia* 24, 96–103. doi: 10.1109/MMUL.2017.38

Iuliani, M., Fontani, M., and Piva, A. (2021). A leak in prnu based source identification–questioning fingerprint uniqueness. *IEEE Access* 9, 52455–52463. doi: 10.1109/ACCESS.2021.3070478

Iuliani, M., Fontani, M., Shullani, D., and Piva, A. (2019). Hybrid reference-based video source identification. *Sensors* 19, 649. doi: 10.3390/s19030649

Lainema, J., Hannuksela, M. M., Vadakital, V. K. M., and Aksu, E. B. (2016). "Hevc still image coding and high efficiency image file format," in *2016 IEEE International Conference on Image Processing (ICIP)* (Phoenix, AZ: IEEE), 71–75.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Liu, Y., Chen, X., Liu, C., and Song, D. (2016). Delving into transferable adversarial examples and black-box attacks. *arXiv preprint arXiv:1611.02770*. doi: 10.48550/arXiv.1611.02770

Lukas, J., Fridrich, J., and Goljan, M. (2006). Digital camera identification from sensor pattern noise. *IEEE Trans. Inf. Forensics Security* 1, 205–214. doi: 10.1109/TIFS.2006.873602

Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., and Swami, A. (2016). Practical black-box attacks against deep learning systems using adversarial examples. *arXiv preprint arXiv:1602.02697* 1, 3. doi: 10.1145/3052973.3053009

Ras, G., van Gerven, M., and Haselager, P. (2018). "Explanation methods in deep learning: users, values, concerns and challenges," in *Explainable and Interpretable Models in Computer Vision and Machine Learning* (Cham: Springer), 19–36.

Ras, G., Xie, N., van Gerven, M., and Doran, D. (2022). Explainable deep learning: a field guide for the uninitiated. *J. Artif. Intell. Res.* 73:329–397. doi: 10.1613/jair.1.13200

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* 115, 211–252. doi: 10.1007/s11263-015-0816-y

Schwab, M., Karrenbach, N., and Claerbout, J. (2000). Making scientific computations reproducible. *Comput. Sci. Eng.* 2, 61–67. doi: 10.1109/5992.881708

Shullani, D., Fontani, M., Iuliani, M., Shaya, O. A., and Piva, A. (2017). Vision: a video and image dataset for source identification. *EURASIP J. Inf. Security* 2017, 1–16. doi: 10.1186/s13635-017-0067-2

Strubell, E., Ganesh, A., and McCallum, A. (2019). Energy and policy considerations for deep learning in nlp. *arXiv preprint arXiv:1906.02243*. doi: 10.18653/v1/P19-1355

Sullivan, G. J., Ohm, J.-R., Han, W.-J., and Wiegand, T. (2012). Overview of the high efficiency video coding (hevc) standard. *IEEE Trans. Circ. Syst. Video Technol.* 22, 1649–1668. doi: 10.1109/TCSVT.2012.2221191

Taubman, D. S., and Marcellin, M. W. (2002). Jpeg2000: standard for interactive imaging. *Proc. IEEE* 90, 1336–1357. doi: 10.1109/JPROC.2002.800725

Vandewalle, P. (2019). "Code availability for image processing papers: a status update," in *WIC IEEE SP Symposium on Information Theory and signal Processing in the Benelux, Date: 2019/05/28-2019/05/29* (Gent).

Vandewalle, P., Kovacevic, J., and Vetterli, M. (2009). Reproducible research in signal processing. *IEEE Signal Process. Mag.* 26, 37–47. doi: 10.1109/MSP.2009.932122

Wiegand, T., Sullivan, G. J., Bjontegaard, G., and Luthra, A. (2003). Overview of the H.264/avc video coding standard. *IEEE Trans. Circ. Syst. Video Technol.* 13, 560–576. doi: 10.1109/TCSVT.2003.815165

Yuan, X., He, P., Zhu, Q., and Li, X. (2019). Adversarial examples: attacks and defenses for deep learning. *IEEE Trans. Neural Netw. Learn. Syst.* 30, 2805–2824. doi: 10.1109/TNNLS.2018.2886017

Zhang, Q.-S., and Zhu, S.-C. (2018). Visual interpretability for deep learning: a survey. *Front. Inf. Technol. Electron. Eng.* 19, 1700808. doi: 10.1631/FITEE.1700808

Zhang, T., and Mao, S. (2019). An overview of emerging video coding standards. *GetMobile Mobile Comp. Comm.* 22, 13–20. doi: 10.1145/3325867.3325873

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.