ACM Journals

CALL FOR PAPERS

ACM Transactions on Multimedia Computing, Communications, and Applications

Special Issue on ***Realistic Synthetic Data: Generation, Learning, Evaluation***

Guest Editors

- **Bogdan Ionescu**, Universitatea Politehnica din București, Romania (bogdan.ionescu@upb.ro)
- **Ioannis Patras**, Queen Mary University of London, UK (i.patras@qmul.ac.uk)
- **Henning Muller**, University of Applied Sciences Western Switzerland, Switzerland (henning.mueller@hevs.ch)
- **Alberto Del Bimbo**, Università degli Studi di Firenze, Italy (alberto.delbimbo@unifi.it)

In the current context of Machine Learning (ML) and Deep Learning (DL), data and especially high-quality data are central for ensuring proper training of the networks. It is well known that DL models require an important quantity of annotated data to be able to reach their full potential. Annotating content for models is traditionally made by human experts or at least by typical users, e.g., via crowdsourcing. This is a tedious task that is time consuming and expensive -- massive resources are required, content has to be curated and so on. Moreover, there are specific domains where data confidentiality makes this process even more challenging, e.g., in the medical domain where patient data cannot be made publicly available, easily.

With the advancement of neural generative models such as Generative Adversarial Networks (GAN), or, recently diffusion models, a promising way of solving or alleviating such problems that are associated with the need for domain specific annotated data is to go toward realistic synthetic data generation. These data are generated by learning specific characteristics of different classes of target data. The advantage is that these networks would allow for infinite variations within those classes while producing realistic outcomes, typically hard to distinguish from the real data. These data have no proprietary or confidentiality restrictions and seem a viable solution to generate new datasets or augment existing ones. Existing results show very promising results for signal generation, images etc.

Nevertheless, there are some limitations that need to be overcome so as to advance the field. For instance, how can one control/manipulate the latent codes of GANs, or the diffusion process, so as to produce in the output the desired classes and the desired variations like real data? In many cases, results are not of high quality and selection should be made by the user, which is like manual annotation. Bias may intervene in the generation process due to the bias in the input dataset. Are the networks trustworthy? Is the generated content violating data privacy? In some cases one can predict based on a generated image the actual data source used for training the network. Would it be possible to train the networks to produce new classes and learn causality of the data? How do we objectively assess the quality of the generated data? These are just a few open research questions.

Topics

In this context, the special issue is seeking innovative algorithms and approaches addressing the following topics (but is not limited to):

- Synthetic data for various modalities, e.g., signals, images, volumes, audio, etc.
- Controllable generation for learning from synthetic data.
- Transfer learning and generalization of models.
- Causality in data generation.
- Addressing bias, limitations, and trustworthiness in data generation.
- Evaluation measures/protocols and benchmarks to assess quality of synthetic content.
- Open synthetic datasets and software tools.
- Ethical aspects of synthetic data.

Important Dates

- Submission deadline: 31 March 2023
- First-round review decisions: 30 June 2023
- Deadline for revised submissions: 31 July 2023
- Notification of final decisions: 30 September 2023
- Tentative publication: December 2023

Submission Information

Prospective authors are invited to submit their manuscripts electronically through the ACM TOMM online submission system (see https://mc.manuscriptcentral.com/tomm) while adhering strictly to the journal guidelines (see https://tomm.acm.org/authors.cfm). For the article type, please select the Special Issue denoted **SI: Realistic Synthetic Data: Generation, Learning, Evaluation**.

Submitted manuscripts should not have been published previously, nor be under consideration for publication elsewhere. If the submission is an extended work of a previously published conference paper, please include the original work and a cover letter describing the new content and results that were added. According to ACM TOMM publication policy, previously published conference papers can be eligible for publication provided that at least 40% new material is included in the journal version.

For questions and further information, please contact **Bogdan Ionescu / bogdan.ionescu@upb.ro.**

Acknowledgements

The Special Issue is endorsed by the AI4Media “A Centre of Excellence delivering next generation AI Research and Training at the service of Media, Society and Democracy” H2020 ICT-48-2020 project https://www.ai4media.eu/.