Translational challenges for synthetic imaging in cardiology

Max L. Olender [1]*, **Farhad R. Nezami** [2], **Lambros S. Athanasiou** [1], **José M. de la Torre Hernández** [3], **and Elazer R. Edelman** [1,4]

[1]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; [2]Thoracic and Cardiac Surgery Division, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; [3]Department of Cardiology, Hospital Universitario Marques de Valdecilla, IDIVAL, Santander, Spain; and [4]Cardiovascular Division, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

We are delighted that Higaki et al. have continued the dialogue initiated by our recent teaching series article[1] about synthetic imaging and prospective clinical uses in cardiology.[2] We agree that forms of artificial intelligence (AI) that construct synthetic images—including generative adversarial networks (GANs)—represent potentially powerful tools in cardiovascular intervention, but that precautions must be taken to ensure their positive contributions to medical practice. Higaki et al.[2] specifically raised concerns regarding image evaluation, our method's scalability, and the risk of technology-mediated misdiagnosis.

The letter highlights an open question regarding how the quality and validity of synthetic images should be evaluated.[2] Image quality and validity are intertwined concepts, though neither attribute guarantees the other. Quality typically reflects a sense of realism arising from image texture, style, and self-consistency; validity usually indicates whether intended information is conveyed or the outcome is similar to a target (i.e. ground truth) image. Higaki et al.[2] note that there is no absolute measure to evaluate synthetic image quality or validity, and indeed there can be no universal, one-size-fits-all criteria. At least 34 assorted metrics of image quality have been employed to train or evaluate GANs in medical imaging applications.[3] The approach to medical image generation we described[1] was trained with two of these to optimize both image quality and validity—through supervised learning—using a common, effective combination of image and adversarial losses.[4] Notably, synthesized images could be directly compared with the target images they emulated to confirm whether each aim was achieved.

However, for the proposed applications, image realism was not an intrinsic aim. While realism is valuable and potentially sufficient for educational use or dataset augmentation,[3] such a standard is inadequate—potentially counterproductive—for evaluating enhanced or augmented images. Benefits of the introduced framework include the integration of additional information or processing algorithms to exceed real images. This can lead, for example, to the desirable—but

unrealistic—absence of obscured regions observed in real optical coherence tomography (OCT) images.[1]

Ultimately, the most important feature of any image—synthetic or real—is the ability to convey information in usable form. As we noted when explaining the fundamentals of morphology-based image generation,[1] future work towards clinical translation will require robust examination of the functional validity and value of these images. Evaluation should be application-driven in the context of outcomes, with images judged by their ability to replace or exceed real data in the performance of analytical tasks.[3,5] Furthermore, hierarchical evaluation frameworks, ranging from technical efficacy (i.e. image quality) through societal efficacy (i.e. cost-effectiveness), have long been espoused for medical imaging technologies.[6]

Given the ongoing evolution of imaging technologies and use trends, generalizability is key to the sustained viability of medical image synthesis in cardiology. Regarding the question of scalability,[2] our approach is particularly promising. Matched pairs of images from different modalities are not required for training the AI, only images matched with corresponding morphology.[1] As such, image generation on the basis of morphology allows for the integration of any number and type of information sources that elucidate tissue distribution, unlike direct image translation, e.g. through supervised image synthesis. Thus, the morphology-mediated approach potentially offers greater flexibility, ability to integrate domain knowledge, and scalability, while also facilitating expanded utility (e.g. through the integration of patient-specific models).

There are key limitations that should be noted, however. The first is that full morphological maps of synthetically imaged tissue must be provided to our method's conditional GAN. For instance, coronary angioscopy (CAS), yielding minimal information on plaque location and composition, would be insufficient to generate comprehensive OCT or intravascular ultrasound (IVUS) images, which convey mural structure. However, CAS could potentially augment other modalities which provide mural structure, thus refining the morphological input

map used to generate synthetic OCT or IVUS images. Similarly, images generated in the style of a modality do not inherit its resolution or accuracy, but rather retain that of the source(s) of rendered morphology. As a practical example, if virtual histology IVUS (VH-IVUS) were used to generate synthetic OCT, an identified thin-cap fibroatheroma (TCFA) should be diagnosed as VH-IVUS-defined TCFA—not OCT-derived TCFA.

These shortcomings and challenges ultimately culminate in the risk of misdiagnosis caused by AI-generated synthetic images. As Higaki et al.[2] imply, reliable conveyance of accurate, unadulterated information is indispensable to the success of this technology. Eye-opening work compellingly illustrated how AI can falter in this aim—GANs trained with biased data artificially introduced or eliminated key pathological features.[7] (Such results align with our own previous work, which showed how training data distribution can bias intravascular image segmentation.[8]) While the training method cautioned against in the aforementioned work[7] fundamentally differs from our own,[1] use of varied datasets representative of intended patient populations is critical,[9] and generators should not be relied upon to synthesize images of scenarios beyond the scope of their training dataset.[7]

Further safeguards are present in our framework, though more are warranted. A benefit of morphology-mediated image generation is its inherent auditability; underlying morphology which the image intends to convey can always be checked, as can source images which yielded that tissue distribution. Furthermore, the human-in-the-loop central to our core aim—to facilitate, rather than automate, image interpretation—provides oversight. However, to facilitate clinical translation, such systems should also transparently convey uncertainty.[9,10] Conflicts between information sources or uncertainty in tissue segmentation, particularly from automated classifiers, should be propagated. Users should be informed accordingly by accompanying or overlaid labels of any segments of the image which are less credible, allowing clinicians to exercise appropriate caution while inspiring trust in the system.[10]

As emphasized here, in our former work,[1] and by Higaki et al.,[2] critical translational challenges, including robust and proper evaluation and validation, scalability, and mitigation of misdiagnosis risk, need to be overcome before synthetic imaging can realize its promise in the clinic. However, these challenges are not insurmountable and should not preclude or deter further work in the field. We continue to look forward to the time when medical images generated by AI—developed with caution and thoughtfulness—augment the cardiologist's visual clinical workflow and enhance clinical practice.

Lead author biography

Max L. Olender is a postdoctoral associate in the Harvard-MIT Biomedical Engineering Center at the Massachusetts Institute of Technology (Cambridge, MA, USA), where he completed a PhD (2021) in mechanical engineering. He also holds B.S.E. (2015) and M.S.E. (2016) degrees in mechanical and biomedical engineering from the University of Michigan (Ann Arbor, MI, USA). He previously conducted research at the University of Michigan with the Biomechanics Research Laboratory and the Neuromuscular Lab. Dr Olender is a member of the Biophysical Society, European Society of Biomechanics, Institute of Electrical and Electronics Engineers, and American Association for the Advancement of Science.

References

1. Olender ML, de la Torre Hernández JM, Athanasiou LS, Nezami FR, Edelman ER. Artificial intelligence to generate medical images: augmenting the cardiologist's visual clinical workflow. *Eur Heart J Digit Health* 2021;**2**:539–544.

2. Higaki A, Miyoshi T, Yamaguchi O. Concerns in the use of adversarial learning for image synthesis in cardiovascular intervention. *Eur Heart J Digit Health* 2021;**2**: 556–556.

3. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. *Med Image Anal* 2019;**58**:101552.

4. Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Honolulu, HI; 2017. p5967–5976. doi: 10.1109/CVPR.2017.632.

5. Frangi AF, Tsaftaris SA, Prince JL. Simulation and synthesis in medical imaging. *IEEE Trans Med Imaging* 2018;**37**:673–679.

6. Thornbury JR. Eugene W. Caldwell Lecture. Clinical efficacy of diagnostic imaging: love it or leave it. *Am J Roentgenol* 1994;**162**:1–8.

7. Cohen JP, Luck M, Honari S. Distribution matching losses can hallucinate features in medical image translation. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, and Fichtinger G (eds). *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Cham: Springer; 2018. p529–536. doi: 10.1007/978-3-030-00928-1_60.

8. Gowrishankar A, Athanasiou L, Olender M, Edelman E. Neural network training data profoundly impacts texture-based intravascular image segmentation. In: *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE, Athens, Greece; 2019. p989–993. doi: 10.1109/BIBE.2019.00184.

9. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;**25**:44–56.

10. Tomsett R, Preece A, Braines D, Cerutti F, Chakraborty S, Srivastava M, Pearson G, Kaplan L. Rapid trust calibration through interpretable and uncertainty-aware AI. *Patterns* 2020;**1**:100049.