KEYNOTE TALK

Monday, October 3, 2022 at 9:00am

Towards Scaling Up GANs

Eli Shechtman
Adobe Research, USA

Abstract: Generative adversarial networks (GANs) have progressed tremendously since their introduction in 2014. They can generate high-quality imagery and their latent space lends itself to editing real images in an intuitive and controllable way. However, they are known to have limitations related to their scalability. They work well when trained on datasets of a single object category, but struggle with more complex scenes. GANs are also limited in the resolution of images they can generate and train on, typically showing results up to 1K pixel resolution that push the current hardware to the limits in memory and training time. To address these, I will first describe a mid-level image representation for a generative model of scenes. The representation is mid-level in that it is neither per-pixel nor per-image; rather, scenes are modeled as a collection of spatial, depth-ordered "blobs" of features. When trained on scenes, our model learns to associate different blobs with different entities in the scene and to arrange these blobs to capture scene layout. We demonstrate this emergent behavior by showing that, despite training without any supervision, our method enables applications such as easy manipulation of objects within a scene and scales well to a diverse dataset of multiple scene categories. I will then describe 'any-resolution' training of GANs that can exploit the variety of image resolutions available in the wild, learning from pixels that are usually discarded, to enable high- and continuously-variable resolution synthesis. We achieve this by switching from the common fixed-resolution thinking, to a novel 'any-resolution' approach, where the original size of each training image is preserved. We introduce a new class of generators that can learn from this multi-resolution signal to synthesize images at any resolution, and show how to train them by sampling patches at multiple scales. Our experiments show generated images from several categories with both coherent global layouts and realistic local details, going beyond 2K and up to 8K resolution. Finally, I will relate these scalability efforts to other recent large-scale generative models (such as
Dall-E 2, Imagen and others).

Speaker Bio-Sketch: Dr. Eli Shechtman is a Senior Principal Scientist at Adobe Research. He received his B.Sc. in Electrical Engineering from Tel-Aviv University in 1996 and his M.Sc. and Ph.D. in Applied Math and Computer Science from the Weizmann Institute of Science in 2003 and 2007. He then joined Adobe and also shared his time as a post-doc at the University of Washington between 2007-2010. His research interests include computer vision, computer graphics and machine learning. In particular, he is focusing on generative modeling and editing of visual data. He has published over 100 academic publications, most of them in top venues and journals in computer vision, graphics and machine learning. He served as a Technical Paper Committee member at SIGGRAPH 2013/14, was an Area Chair at CVPR 2015/17/20, ICCV 2015/19/21, ECCV 2022 and was an Associate Editor for TPAMI from 2016 to 2020. He has received several honors and awards, including the Best Paper prize at ECCV 2002, a Best Paper award at WACV 2018 and Helmholtz "Test of Time" prize at ICCV 2017. Two of his papers were chosen to be published as "Research Highlight" papers in the Communication of the ACM (CACM) journal. Some of his research can be found in Adobe's products such as Photoshop's Content Aware Fill, Smart Portrait and Landscape Mixer, Content Aware Fill for Video in After Effects, Upright in Lightroom and Characterizer in Character Animator.