Tech giants assemble in the first virtual Computer Vision and Pattern Recognition Conference 2020

OROBI BAKHTIAR

CVPR is the premier annual Computer Vision and Pattern Recognition Conference. With first-in-class technical content, the main program, tutorials, workshops, a leading-edge expo, and attended by more than 9,000 people annually, CVPR creates a one-of-a-kind opportunity for networking, recruiting, inspiration, and motivation. CVPR 2020, originally scheduled to take place 14-19 June 2020 at the Washington State Convention Center in Seattle Washington, has been a fully virtual event this year. Authors and presenters have virtually-delivered presentations and have planned to engage in live Q&A with attendees.

As the largest conference covering every aspect of computer vision and pattern recognition, machine learning, and artificial intelligence, CVPR features more than 1,500 presentations from industry leaders, including Amazon Web Services, Alibaba Group, Apple, Google, Microsoft, Waymo and many others. In exclusive "fireside chat" interview sessions with Microsoft CEO Satya Nadella and Amazon Web Services Senior Vice President Charlie Bell, CVPR attendees will hear first-hand how the tech giants continue to fuel advancements in AI technologies and upcoming developments.

"CVPR draws together the leading technologists exploring AI and machine learning and presents foundational research driving new opportunities. The dialogue at CVPR between researchers and industry leaders helps spur the next round of scientific innovation", remarked Ramin Zabih, Cornell Professor of Computer Science and Co-Chair of the CVPR 2020 Organizing Committees.

Some of the research papers that started trending within the AI research community months before their actual presentation at CVPR 2020. These papers cover the efficiency of object detectors, novel techniques for converting RGB-D images into 3D photography, and autoencoders that go beyond the capabilities of generative adversarial networks (GANs) with respect to image generation and manipulation.

EfficientDet: Scalable and Efficient Object Detection

The large size of object detection models deters their deployment in real-world applications such as self-driving cars and robotics. To address this problem, the Google Research team introduced two optimizations, namely (1) a weighted bi-directional feature pyramid network (BiFPN) for efficient multi-scale feature fusion and (2) a novel compound scaling method. By combining these optimizations with the EfficientNet backbones, the authors developed a family of object detectors, called EfficientDet. The experiments demonstrated that these object detectors consistently achieve higher accuracy with far fewer parameters and multiply-adds (FLOPs).

The high accuracy and efficiency of the EfficientDet detectors may enable their application for real-world tasks, including self-driving cars and robotics.

3D Photography using Context-aware Layered Depth Inpainting

The research team presented a new learning-based approach to generating a 3D photo from a single RGB-D image. The depth in the input image can either come from a cell phone with a stereo camera or be estimated from an RGB image. The authors suggested explicitly storing connectivity across pixels in the representation. To deal with the resulting complexity of the topology and the difficulty of applying a global CNN to the problem, the research team broke the problem into many local inpainting sub-problems that are solved iteratively. The introduced algorithm results in 3D photos with synthesized textures and structures in occluded regions. The experiments demonstrated its effectiveness compared to the existing state-of-the-art techniques.

3D photography provides a much more immersive experience than usual 2D images, so the ability to easily generate a 3D photo from a single RGB-D image can be useful in many business areas, including real estate, e-commerce, marketing, and advertising.

Adversarial Latent Autoencoders

The research group from West Virginia University investigated if autoencoders can have the same generative power as GANs while learning disentangled representation. In particular, they introduced an autoencoder, called Adversarial Latent Autoencoder (ALAE), that can generate images with quality comparable to state-of-the-art GANs while also learning a less entangled representation. This is achieved by allowing the latent distribution to be learned from data and the output data distribution to be learned with an adversarial strategy. Finally, the autoencoder's reciprocity is imposed in the latent space. The experiments demonstrated that the introduced autoencoder architecture with the generator derived from a StyleGAN, called StyleALAE, has generative power comparable to that of StyleGAN but can also produce face reconstructions and image manipulations based on real images rather than generated.

The suggested approach enables images to be generated and manipulated with a high level of visual detail, and thus may have numerous applications in real estate, marketing and advertising.