Automatic image retouching on your phone

August 2 2017, by Larry Hardesty

Researchers describe a new system that can automatically retouch images in the style of a professional photographer. It can run on a cellphone and it's so fast that it can display retouched images in real-time, so that the photographer can see the final version of the image while still framing. Credit: Massachusetts Institute of Technology

The data captured by today's digital cameras is often treated as the raw

material of a final image. Before uploading pictures to social networking sites, even casual cellphone photographers might spend a minute or two balancing color and tuning contrast, with one of the many popular image-processing programs now available.

This week at Siggraph, the premier digital graphics conference, researchers from MIT's Computer Science and Artificial Intelligence Laboratory and Google are presenting a new system that can automatically retouch images in the style of a professional photographer. It's so energy-efficient, however, that it can run on a cellphone, and it's so fast that it can display retouched images in real-time, so that the photographer can see the final version of the image while still framing the shot.

The same system can also speed up existing image-processing algorithms. In tests involving a new Google algorithm for producing high-dynamic-range images, which capture subtleties of color lost in standard digital images, the new system produced results that were visually indistinguishable from those of the algorithm in about one-tenth the time—again, fast enough for real-time display.

The system is a machine-learning system, meaning that it learns to perform tasks by analyzing training data; in this case, for each new task it learned, it was trained on thousands of pairs of images, raw and retouched.

The work builds on an earlier project from the MIT researchers, in which a cellphone would send a low-resolution version of an image to a web server. The server would send back a "transform recipe" that could be used to retouch the high-resolution version of the image on the phone, reducing bandwidth consumption.

"Google heard about the work I'd done on the transform recipe," says

Michaël Gharbi, an MIT graduate student in electrical engineering and computer science and first author on both papers. "They themselves did a follow-up on that, so we met and merged the two approaches. The idea was to do everything we were doing before but, instead of having to process everything on the cloud, to learn it. And the first goal of learning it was to speed it up."

Short cuts

In the new work, the bulk of the image processing is performed on a low-resolution image, which drastically reduces time and energy consumption. But this introduces a new difficulty, because the color values of the individual pixels in the high-res image have to be inferred from the much coarser output of the machine-learning system.

In the past, researchers have attempted to use machine learning to learn how to "upsample" a low-res image, or increase its resolution by guessing the values of the omitted pixels. During training, the input to the system is a low-res image, and the output is a high-res image. But this doesn't work well in practice; the low-res image just leaves out too much data.

Gharbi and his colleagues—MIT professor of [electrical engineering](link) and computer science Frédo Durand and Jiawen Chen, Jon Barron, and Sam Hasinoff of Google—address this problem with two clever tricks. The first is that the output of their machine-learning system is not an image; rather, it's a set of simple formulae for modifying the colors of image pixels. During training, the performance of the system is judged according to how well the output formulae, when applied to the original image, approximate the retouched version.

Taking bearings

The second trick is a technique for determining how to apply those formulae to individual pixels in the high-res image. The output of the researchers' system is a three-dimensional grid, 16 by 16 by 8. The 16-by-16 faces of the grid correspond to pixel locations in the source image; the eight layers stacked on top of them correspond to different pixel intensities. Each cell of the grid contains formulae that determine modifications of the color values of the source images.

That means that each cell of one of the grid's 16-by-16 faces has to stand in for thousands of pixels in the high-res image. But suppose that each set of formulae corresponds to a single location at the center of its cell. Then any given high-res pixel falls within a square defined by four sets of formulae.

Roughly speaking, the modification of that pixel's color value is a combination of the formulae at the square's corners, weighted according to distance. A similar weighting occurs in the third dimension of the grid, the one corresponding to pixel intensity.

The researchers trained their system on a data set created by Durand's group and Adobe Systems, the creators of Photoshop. The data set includes 5,000 images, each retouched by five different photographers. They also trained their system on thousands of pairs of images produced by the application of particular image-processing algorithms, such as the one for creating high-dynamic-range (HDR) images. The software for performing each modification takes up about as much space in memory as a single digital photo, so in principle, a cellphone could be equipped to process images in a range of styles.

Finally, the researchers compared their system's performance to that of a machine-learning system that processed images at full resolution rather than low resolution. During processing, the full-res version needed about 12 gigabytes of memory to execute its operations; the researchers'

version needed about 100 megabytes, or one-hundredth as much. The full-resolution version of the HDR system took about 10 times as long to produce an image as the original algorithm, or 100 times as long as the researchers' [system](#).

"This technology has the potential to be very useful for real-time image enhancement on mobile platforms," says Barron. "Using machine learning for computational photography is an exciting prospect but is limited by the severe computational and power constraints of mobile phones. This paper may provide us with a way to sidestep these issues and produce new, compelling, real-time photographic experiences without draining your battery or giving you a laggy viewfinder experience."

  More information: Paper: "Deep bilateral learning for real-time image enhancement" [groups.csail.mit.edu/graphics/ …](#) [rnet/data/hdrnet.pdf](#)

This story is republished courtesy of MIT News ([web.mit.edu/newsoffice/](#)), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology