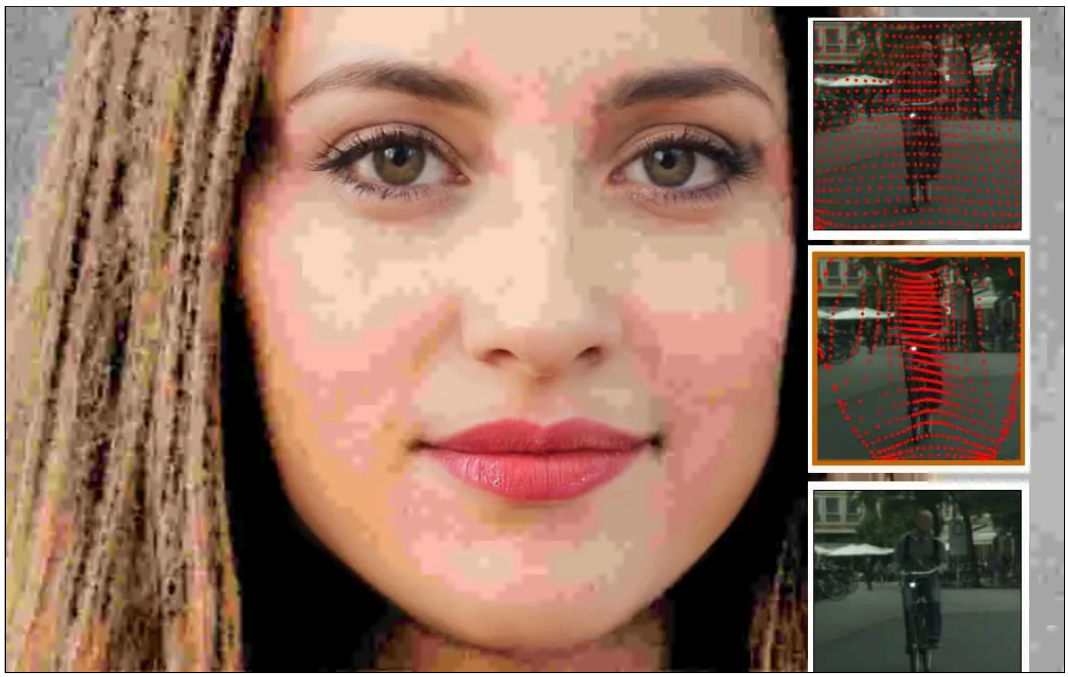
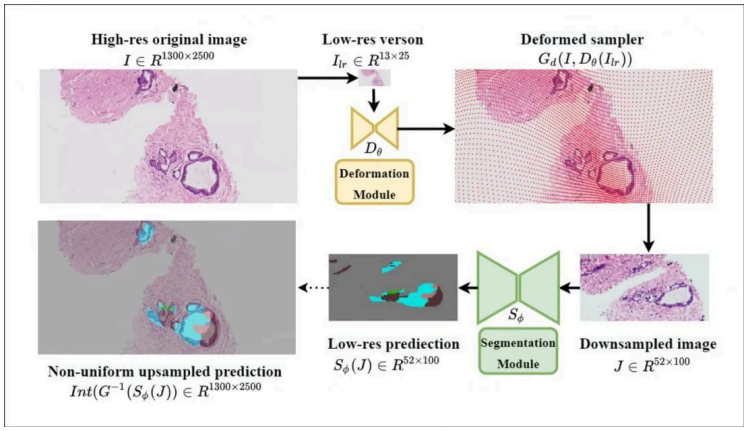
Downsizing High Resolution Images With Machine Learning

Originally published at https://www.unite.ai/downsizing-high-resolution-images-with-machine-learning/



New research from the UK has proposed an improved machine learning method to resize images, based on the perceived value of the various parts of the image content, instead of indiscriminately reducing the dimension (and therefore the quality and extractable features) for all the pixels in the image.

As part of a growing interest in AI-driven compression systems, it's an approach that could eventually inform new codecs for general image compression, though the work is motivated by health imaging, where arbitrary downsampling of high-resolution medical images could lead to the loss of life-saving information.



Representational architecture of the new system. The interstitial deformation module produces a deformation map that corresponds to areas of interest in the imag and direction of the red dots indicate these areas. The map is used not only to downsample, but to reconstruct the primary-interest areas when the image content re-upscaled at the other side of the training process. Source: https://arxiv.org/pdf/2109.11071.pdf

The system applies semantic segmentation to the images – broad blocks, represented as color blocks in the image above, that encompass recognized entities inside the picture, such as *'road'*, *'bike'*, *'lesion'*, et al. The disposition of the semantic segmentation maps are then used to calculate which parts of the photo should not be excessively downsampled.

Entitled *Learning to Downsample for Segmentation of Ultra-High Resolution Images*, the new paper is a collaboration between researchers from the Centre for Medical Image Computing at University College London and researchers from the Healthcare Intelligence department at Microsoft Cambridge.

The (Fairly) Low-Res World of Computer Vision Training

The training of computer vision systems is significantly constrained by the capacity of GPUs. Datasets may contain many thousands of images from which features need to be extracted, but even industrial-scope GPUs tend to peak at 24gb of VRAM, with ongoing shortages affecting availability and cost.

This means that data must be fed through the limited Tensor cores of the GPU in manageable batches, with 8-16 images typical of many computer vision training workflows.

There are not many obvious solutions: even if VRAM were unlimited and CPU architectures could accommodate that kind of throughput from the GPU without forming an architectural bottleneck, very high batch sizes will tend to derive high-level features at the expense of the more detailed transformations that may be critical to the usefulness of the final algorithm.

Increasing the resolution of the input images will mean that you have to use smaller batch sizes to fit data in the 'latent space' of the GPU training. This, conversely, is likely to produce a model that is 'eccentric' and overfitted.

Neither does adding extra GPUs help, at least in the most common architectures: while multiple-GPU setups can speed up training times, they can also compromise the integrity of training results, like two adjacent factories working on the same product, with only a phone line to coordinate their efforts.

Intelligently Resized Images

What's left is that the most relevant sections of a typical image for a computer vision dataset could, with the new method, be preserved intact in the automatic resizing that occurs when very high-resolution images must be downscaled to fit an ML pipeline.

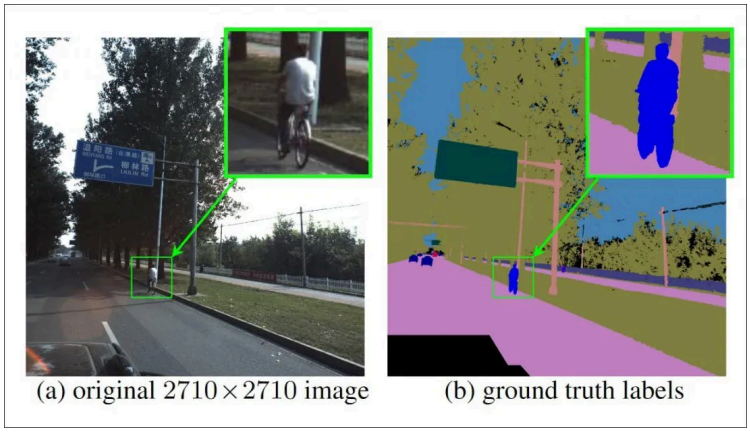
This is a separate challenge to the problem of lossy artefacts in machine learning datasets, where quality is lost in automated resizing pipelines because the compression codec throws away too much (usually irrecoverable) information.

Rather, in this case, even saving to a lossless image format (such as PNG with LZW compression) cannot recover the information that's generically discarded when resizing (for instance) a Magnetic Resonance Imaging (MRI) scan down from often record-breaking dimensions to a more credible typical 256×256 or 512×512 pixels resolution.

To make things worse, depending on the requirements of the framework, black borders will often be added to rectangular source images as a routine data processing task, in order to produce a genuinely square input format for neural network processing, further reducing available space for potentially crucial data.

The researchers from UCL and Microsoft instead propose to make the resizing process more intelligent, effectively using what has always been a generic stage in the pipeline to highlight areas of interest, off-loading some of the interpretive burden from the machine learning system through which the images will ultimately pass.

The method, the researchers claim, improves on a 2019 offering (image below) which sought similar gains by focusing quality-attention at the *boundaries* of objects.



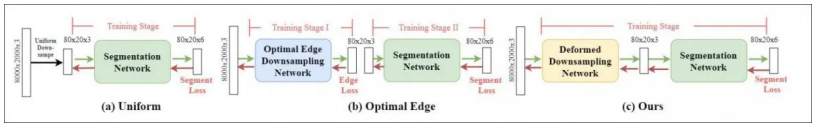
From 'Efficient Segmentation: Learning Downsampling Near Semantic Boundaries', Marin et al., 2019. Source: https://arxiv.org/pdf/1907.07156.pdf

As the new work notes, this approach assumes that areas of interest gather at boundaries, whereas examples from medical imaging, such as annotated cancer regions, depend on higher-level context, and may appear as easily-discarded details within broader areas in an image, rather than at edges.

Learnable Downsampler

The new research proposes a *learnable downsampler* called a deformation module, that's jointly trained with a parallel segmentation module, and can therefore be informed about areas of interest identified by semantic segmentation, and prioritize these during the downsampling process.

The authors tested the system on several popular datasets, including Cityscapes, DeepGlobe and a local Prostate Cancer Histology dataset, 'PCa-Histo'.



Three approaches: on the left, existing 'uniform' downsampling; in the middle, the 'optimal edge' approach from the 2019 paper; on the right, the architecture behir entity recognition in a semantic segmentation layer.

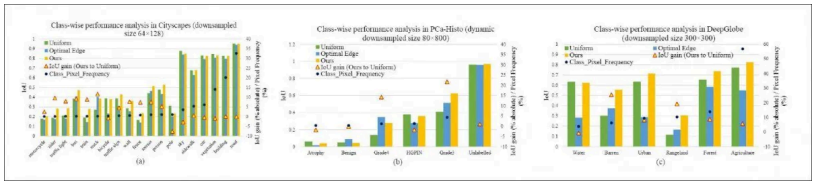
A similar approach has been tried for a classifier proposed in 2019, but the authors of the current paper contend that this method does not adequately regularize the areas of emphasis, potentially missing vital areas in a medical imaging context.

Results

The deformation module in the new system is a small Convolutional Neural Network (CNN), while the segmentation layer is a deep CNN architecture employing HRNetV2-W48. Pyramid Scene Parsing Network (PSP-net) was used as a sanity check layer for the CityScapes tests.

The aforementioned datasets were tested with the new framework, using uniform resampling (the customary method), the optimal edge method from 2019, and the new approach's leveraging of semantic segmentation.

The authors report that the new method shows *'clear advantage on identifying and distinguishing the most clinically important classes'*, with an accuracy boost of 15-20%. They further observe that the distance between these classes is often defined as 'the threshold from healthy to cancer'.



Class-wise intersection over union (IoU) analysis across the three methods: left, standard resampling; middle, optimal edge; and right, the new approach.CityScap 64 x 128, with PCaHisto down to 80 x 800, and DeepGlobe down to 300 pixels square.

The report states that their method *'can learn a downsampling strategy, better preserve information and enable a better trade-off.', concluding that* the new framework *'can efficiently learn where to "invest" the limited budget of pixels at downsampling to achieve highest overall return in segmentation accuracy'*.

The main image for this feature's article was sourced from thispersondoesnotexist.com. Updated 3:35pm GMT+2 for text error.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Portfolio site: martinanderson.ai
Contact: martin@martinanderson.ai

Discover More