Image super-resolution using hierarchical VAEs

Paul Jacques Connetable (s151404), Paolo Zifferero (s200149), Prasad Jagtap (s200109)
Course : 02456 Deep Learning
DTU Compute, Technical University of Denmark

Introduction.

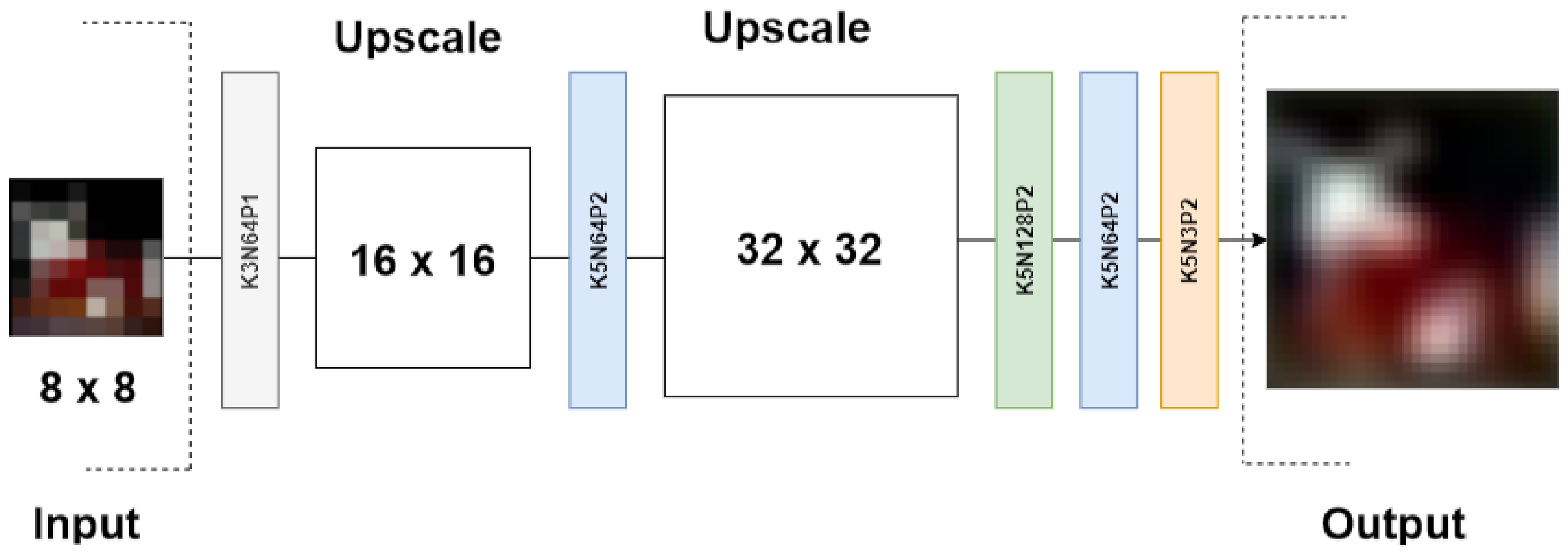
Increasing an image resolution or reconstructing a high-resolution image from a coarse resolution image is a very active area of research. This is called ***image super-resolution (SR).*** However, this can be used to improve the image quality of movies or pictures, as well as store them in coarser resolution resulting to accommodating lesser space. The latest state of the art super-resolution neural network is a deep hierarchical VAE.

The more recent models have proven to be much more adequate to the task, and have also been able to generate new images, separate an image's content & style.

We used the **CIFAR-10** data-set in our project, which consists in a collection of *32 by 32 RGB images*. In order to train the networks on ***SR***, we treat the original images as our high-resolution image targets. We create the low-resolution images by applying a *gaussian filter* and *sub-sampling* the high-resolution images, to obtain *8 by 8 RGB images*. The *gaussian filter* is deemed particularly important as it allows to preserve the information of a pixel by spreading it to the neighboring ones. By doing so, even when performing down sampling the information will not be lost.

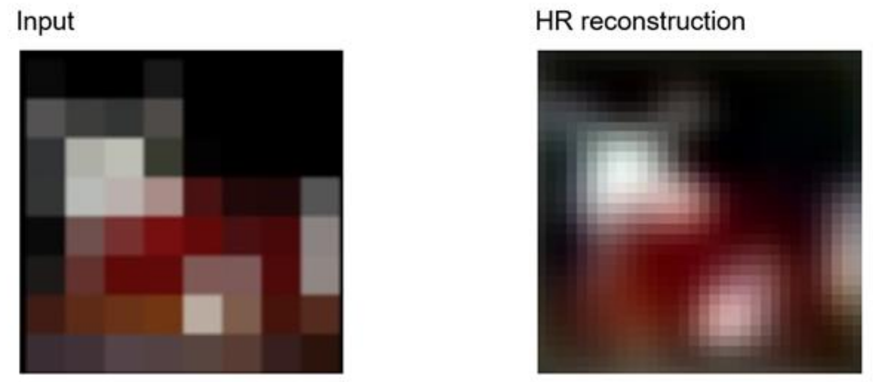
CNN

Convolutional neural networks were first used to perform image ***SR*** in 2015 by C. Dong *et al* [1], proving to have a high potential and flexibility. In the application presented below, the model was inspired by the one presented in the paper, and it has been optimized to improve training times and computational efficiency. The network's input is an *8 by 8 RGB image.* However, upscaling is performed twice to match the target image size.

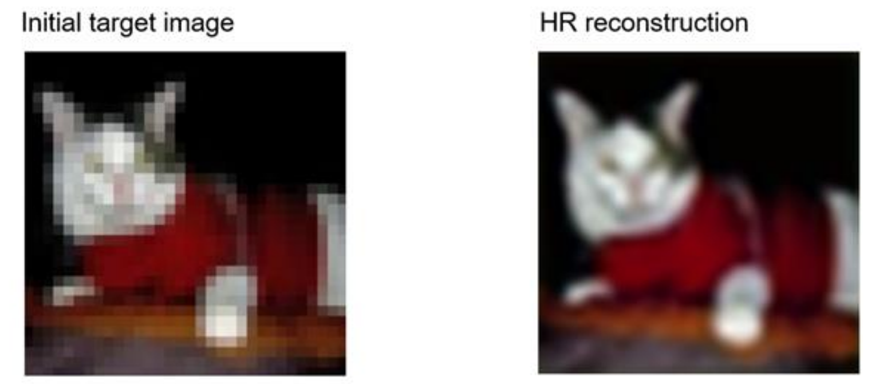


Results

The network is able to create a HR image from a LR one. CIFAR-10 images, however, contain extremely few pieces of information when downscaled with a factor of 4, so the results are barely satisfying;



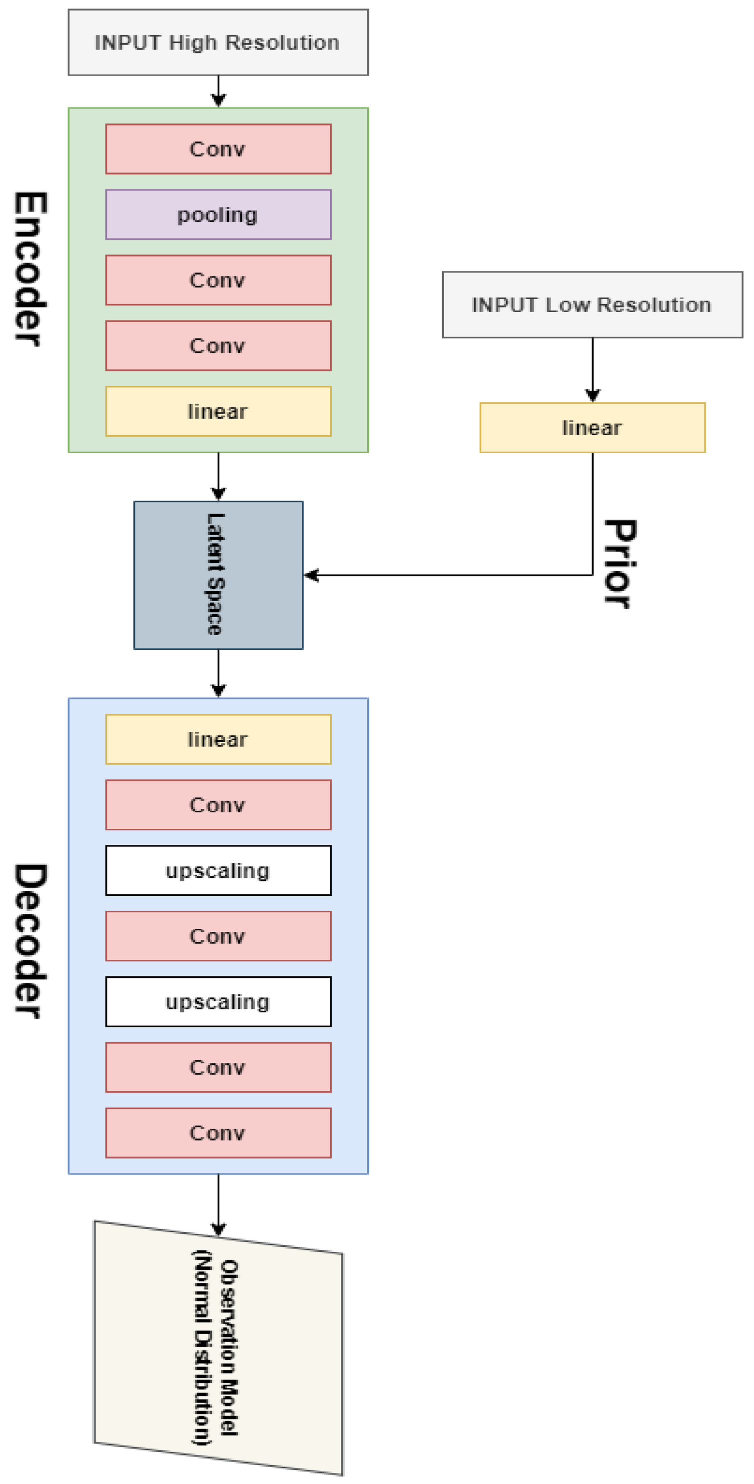
When feeding a bigger image, the improvement in quality is much more appreciable as shown below;



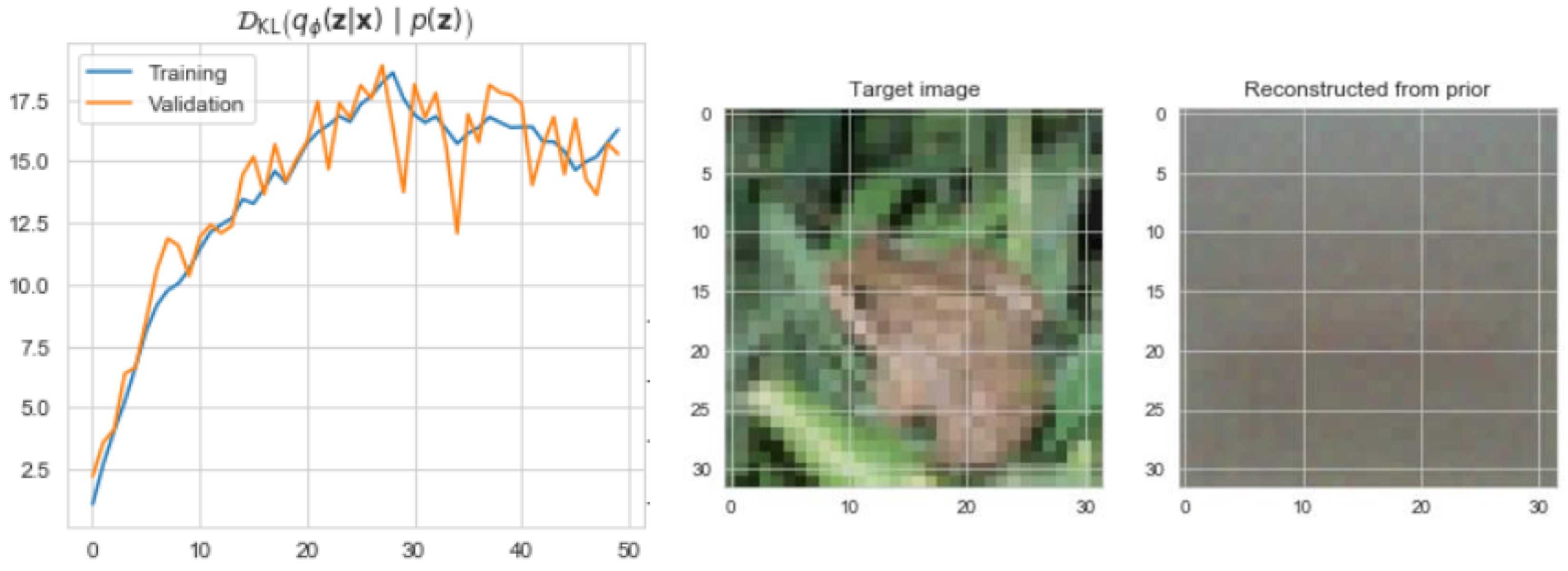
VAE

There are several axes of research based on the use of ***VAEs*** for ***SR***, like in [5], used in a ladder network [2]. They can also be associated with a ***GANs*** like in [3], and the state-of-the art relies on deep hierarchical VAE models [4] like the ***NVAE*** and the ***BIVA*** models.

Before trying the more advanced models, we wanted to implement a simple ***VAE*** used for ***SR***. We used a *latent space* size of ***192***, which corresponds to the total size of our input low resolution images, and our observation model is a normal distribution with a fixed low standard deviation (0.01). Our network architecture is presented on the right. We used ***RELU*** after each step as a non-linearity, except on the last decoder layer, which uses a ***sigmoid*** non-linearity.



Results



The VAE network was very hard to train. Comparatively to the CNN, it needs many epochs to train well, and is longer to train since it is larger. We tried to vary the batch size, the latent space size, the beta parameter the learning rate of the network, and adding a MSE loss to the usual beta elbo loss, but always ended up with mostly grey images. In some cases, we could also end up with bright spots and colors in the reconstruction.

An example of training on 50 epochs is presented, with an example of a test image, and a reconstruction based on its prior. The MSE loss on the whole test set is 0.062.

References

1. C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks", IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 2, pp. 295–307, 2015.
2. C. K. Sønderby, T. Raiko, L. Maaløe, S. K. Sønderby, and O. Winther, "Ladder varia-tional autoencoders", in Advances in neural information processing systems, pp. 3738–3746, 2016.
3. C. Ledig, L. Theis, F. Husz´ar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang et al., "Photo-realistic single image super-resolution using a generative adversarial network," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4681–4690, 2017.
4. A. Vahdat and J. Kautz, "Nvae: A deep hierarchical variational autoencoder", arXivpreprint arXiv:2007.03898, 2020
5. I. Gatopoulos, M. Stol, and J. M. Tomczak, "Super-resolution variational auto-encoders", arXiv preprint arXiv:2006.05218, 2020.