Video resolution upscaling using GAN and U-Net

Adam Formánek

ČVUT - FIT

formaad2@fit.cvut.cz

4. prosince 2021

1 Introduction

Main goal of this assignment is to take short video and increase its resolution using neural networks with Generative Adversarial Network (GAN) and U-Net architecture. These two architectures should be then compared and commented.

2 Input data

Dataset was taken from Kaggle. It is a 5 minute long video of a popular kids show Tom and Jerry.

3 Research

Super resolution is a broad field with a lot of different approaches. Since U-Net and GAN architecture must be used in solution, research was done mainly about these topics.

U-Net is a very common approach for image segmentation. It firstly appeared in [6], where it was used for segmentation of biomedical images. Thanks to the unique architecture, it was then applied for many other purposes, including super resolution.

In [5] was U-Net used for resolution upscaling of x-ray images. Input image is downscaled (encoded) using convolultion and max pool layers. Then by using upsample convolution and skip connections, from the encoding phase containing important information about the original frame, image is reconstructed (decoded) with higher resolution.

Another example can be found in [4], where original architecture from [6] was modified. Firstly all batch normalizations and one convolution layer in each block are removed. Input image is upscaled and has skip connection with the output image. Lastly, error is measured by mixing *Mean Square Error* (MSE) with *weighted Mean Gradient Error* (λ_GMGE), which is proposed for a sharp edge reconstruction.

GAN was first introduced in [1]. This approach consists of two main parts, generator and discriminator. Generator takes some noise on input and generates image, that is passed to the discriminator. Discriminator then measures, how similar the generated image is to the reference image. Good analogy can be visualizing discriminator and generator as oponents, trying to beat each other in minimax game [2].

First implementation of this approach for super resolution, SRGAN, was described in [3]. This approach combines GAN and deep residual network with skip connections. Perceptual loss is used instead of MSE as loss function to improve visual effect of the reconstructed result.

Another approach, found in [7], is a modification of [3]. Upsampling by nearest neighbor is used instead of bicubic interpolation, which is more computationaly efficient, but produces checkerboard artifacts. That is solved by adding convolution layer.

4 Methods

Input dataset have 1280x720 resolution. Preprocessing of the input dataset includes parsing video into individual frames and resizing them to have equal height and width (in particular 512x512). Low and corresponding high resolution images are needed for training the model. So for every frame low resolution one is created by using bicubic downsampling ($4x \rightarrow 128x128$). Original image is then representing high resolution version.

Implementation is not done yet. As for GAN architecture, implementation is currently in progress and will follow [3]. Implementation of U-Net architecture will probably be inspired by [4].

5 Future work

First of all implementation of described approaches must be written and experimentaly tested. Models can be then trained and tested on described dataset and compared to each other.

Reference

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Ge-

nerative adversarial nets [online]. *Advances in neural information processing systems*, 27, 2014.

[2] Jonathan Hui. Medium: Gan - what is generative adversarial networks gan? [online], 2018. [cit. 2021-12-04].

[3] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network [online]. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[4] Zhengyang Lu and Ying Chen. Single image super-resolution based on a modified u-net with mixed gradient loss [online]. *Signal, Image and Video Processing*, pages 1–9, 2021.

[5] Junyoung Park, Donghwi Hwang, Kyeong Yun Kim, Seung Kwan Kang, Yu Kyeong Kim, and Jae Sung Lee. Computed tomography super-resolution using deep convolutional neural network [online]. *Physics in Medicine & Biology*, 63(14):145011, jul 2018. [cit. 2021-12-04].

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation [online], 2015.

[7] Mehdi SM Sajjadi, Bernhard Scholkopf, and Michael Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4491–4500, 2017.