Boosting VFX Production with Deep Learning

Yanir Kleiman
DNEG

Simon Pabst
DNEG

Patrick Nagle
DNEG

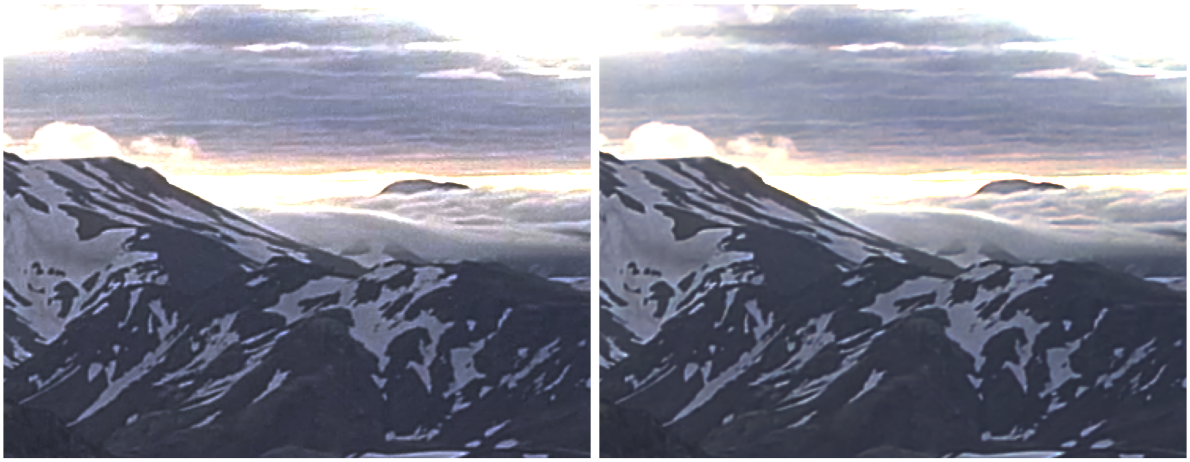

Figure 1: Left: a patch from an input frame taken from a compressed video. Right: the corresponding patch from the restored image. Images are automatically reconstructed using a data driven deep learning approach. Note that noise and compression artifacts such as ringing and blockiness are reduced while details are preserved.

ABSTRACT

Machine learning techniques are not often associated with artistic work such as visual effects production. Nevertheless, these techniques can save a lot of time for artists when used in the right context. In recent years, deep learning techniques have become a widely used tool with powerful frameworks that can be employed in a production environment. We present two deep learning solutions that were integrated into our production pipeline and used in current productions. One method generates high quality images from a compressed video file that contains various compression artifacts. The other quickly locates slates and color charts used for grading in a large set of images. We discuss these particular solutions in the context of previous work, as well as the challenges of integrating a deep learning solution within a VFX production pipeline, from concept to implementation.

KEYWORDS

Deep Learning, Compression Artifacts, Machine Learning

ACM Reference format:

Yanir Kleiman, Simon Pabst, and Patrick Nagle. 2019. Boosting VFX Production with Deep Learning. In *Proceedings of SIGGRAPH '19 Talks, Los Angeles, CA, USA, July 28 - August 01, 2019,* 2 pages.
https://doi.org/10.1145/3306307.3328208

SIGGRAPH '19 Talks, July 28 - August 01, 2019, Los Angeles, CA, USA

ACM ISBN 978-1-4503-6317-4/19/07.
https://doi.org/10.1145/3306307.3328208

1 INTRODUCTION

The work of a skilled artist cannot be replaced by machine learning algorithms. However, these techniques can certainly help artists and relieve them of a lot of menial work. In particular, these methods can assist in handling large volumes of data and low-level image processing. Deep learning methods have pushed the limit of what is possible with machine learning. As a result, it is no longer limited to cutting-edge research but is widely used in many industries. We show how this powerful tool can be used in VFX production, and present two deep learning solutions that were used in our recent shows: one to reconstruct high quality images from a compressed video, and the other to find slates and color charts within images shot on a movie set. These solutions are tailored to our specific needs and more effective than currently available methods.

2 IMAGE RECONSTRUCTION

Visual effects production is a collaborative effort, and at DNEG we often receive data from clients or other production houses. Unfortunately, sometimes this data is not available in a high quality form but only as a compressed movie file which contains various MPEG and JPEG compression artifacts. Our artists then apply various noise reduction techniques and a lot of tedious manual work in order to generate high quality content from low quality frames.

The goal of our deep learning solution is not only to reduce the noise in the image sequence, but to reconstruct image elements that might be lost due to compression: repeated patterns which are broken, sharp edges that become blurry, colors which become desaturated or over saturated, etc, as can be seen in Figure 2.

Note that traditional methods of noise reduction focus on removing different frequencies of noise in a digital or film image.

They cannot reproduce lost elements in the image such as repeated structures, sharp details, or colors, since they lack the high level context that is provided to the neural network.

2.1 Implementation

Deep learning solutions intended for 4K resolutions typically break the image into smaller patches of around 256x256 pixels. This is done in order to fit the model in the GPU memory, where convolutional networks are most efficiently trained. Rather than generating a 256x256 output image for each patch, we generate a central patch of size 24x24, and compare the output to the central patch in the target image. The larger patch provides context to the central patch, which is important to identify repeated patterns and global properties of the image such as color palette and type of content. On the other hand, since the context patch is not part of the loss function, the reconstruction of the central patch can be more accurate. In our experiments, working with small patches provided a significant boost in reconstruction accuracy.

To train the network, we gathered a small collection of reference videos of high quality, and compressed them in a similar manner to the typical compressed input videos. Since the network is patch based, a small collection of 4K frames is enough to produce almost endless patch training data, as each image contains millions of different patches. Figure 2 shows some of the validation samples and their corresponding output. Figure 1 shows the output of our network on a larger image by stitching many small patches.

3 SLATE AND COLOR CHART DETECTION

To facilitate the artists work, we take detailed reference images of everything on the set - from actors, to props, to the location of each shot. For each set of images we shoot a slate, to identify the element, and a color calibration chart, to ensure the element can be graded correctly. These datasets often contain thousands of images, and it is important to tag the slate and color chart images for further processing. We employ deep learning to automatically identify the slate and color chart images in each set, which would otherwise be a time-consuming manual task.

Previously proposed methods for color chart detection [Baumann 2015] often assume a color chart exists in the image and detect the position of individual color patches. Others are not general enough to handle perspective changes, fisheye lenses, etc. We are not aware of any existing methods for detecting slates.

We train a convolutional neural network which outputs a single output between 0 and 1 which indicates whether a slate is present in the image or not. We train a separate network for the color charts. For training data we gathered roughly 10,000 images from previous projects which were already tagged as containing a slate, a color chart, both, or none. Using this data, we experimented with various network architectures and training parameters and were eventually able to reach a very high accuracy rate of about 99% correct predictions on the validation set.

Similar networks can perform different tasks based on their training data. For example, we use the same network structure to identify slates and color charts. Similarly, the image reconstruction network that is trained to reduce compression artifacts can also be trained for other tasks such as super resolution or image completion.

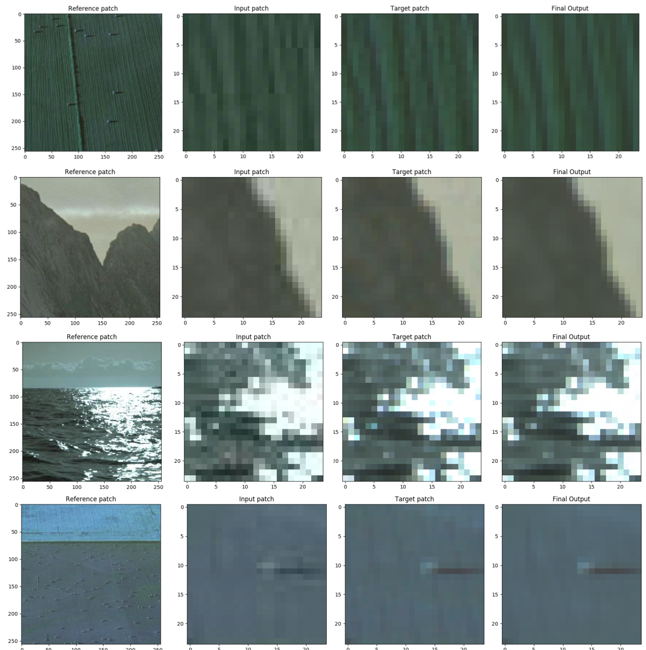

Figure 2: Validation set samples. Each row displays a reference patch which surrounds the input patch on the left, followed by the input patch, the training target to reconstruct, and the final output reconstructed by the network. Our network is able to reconstruct patterns based on the surrounding context and remove blockiness artifacts (first row), reconstruct a sharp edge and remove ringing effects (second row), reconstruct colors which are lost due to compression (first and third row), and reconstruct a shape affected by compression noise (fourth row).

4 FROM RESEARCH TO PRODUCTION

Deep learning methods are a natural fit for the visual effects industry, as they require a lot of data, something which VFX production houses have in abundance: high definition HDR images; before and after shots; high quality renders and 3D models; photogrammetry data; all tagged and labeled in an orderly fashion. Moreover, artist workstations are already fitted with powerful GPUs, which are required to train and run deep learning algorithms efficiently.

In academic research, the goal is often for the method to be successful on a few specific datasets. On the other hand, production tools are expected to work reliably for any given data, even if some degree of error is acceptable. Deep learning solutions should be employed strategically and allow artists to save time and follow through in failure cases. For example, after employing our image reconstruction network, artists can further reduce artifacts by using traditional methods where necessary. This way, we can reduce the load of repetitive work and enable artists to focus on the more creative tasks in visual effects production.

REFERENCES

Ryan Baumann. 2015. Automatic ColorChecker Detection, a Survey. (2015). https://ryanfb.github.io/etc/2015/07/08/automatic_colorchecker_detection.html