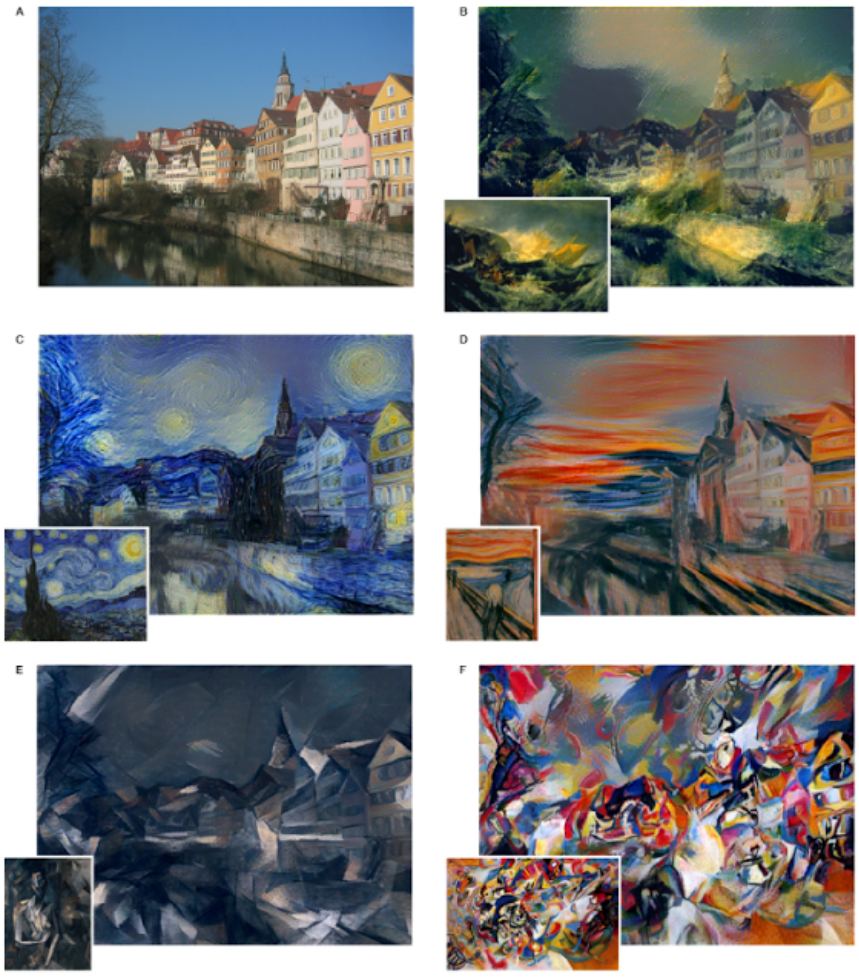
FPGram Project Design

Tonye Brown, tb2553
Diana Valverde, drv2110

Aim

The general premise of our project is to create a photo booth application with the ability to use a pre-set filter to render images in the style of well-known artists. By using convolutional neural networks, we can extract high-level feature data from user images and low-level features of artistic paintings to serve as the constraints of the construction of a new image. This is a technique developed by Gatys et al. which has been used to create the following images:

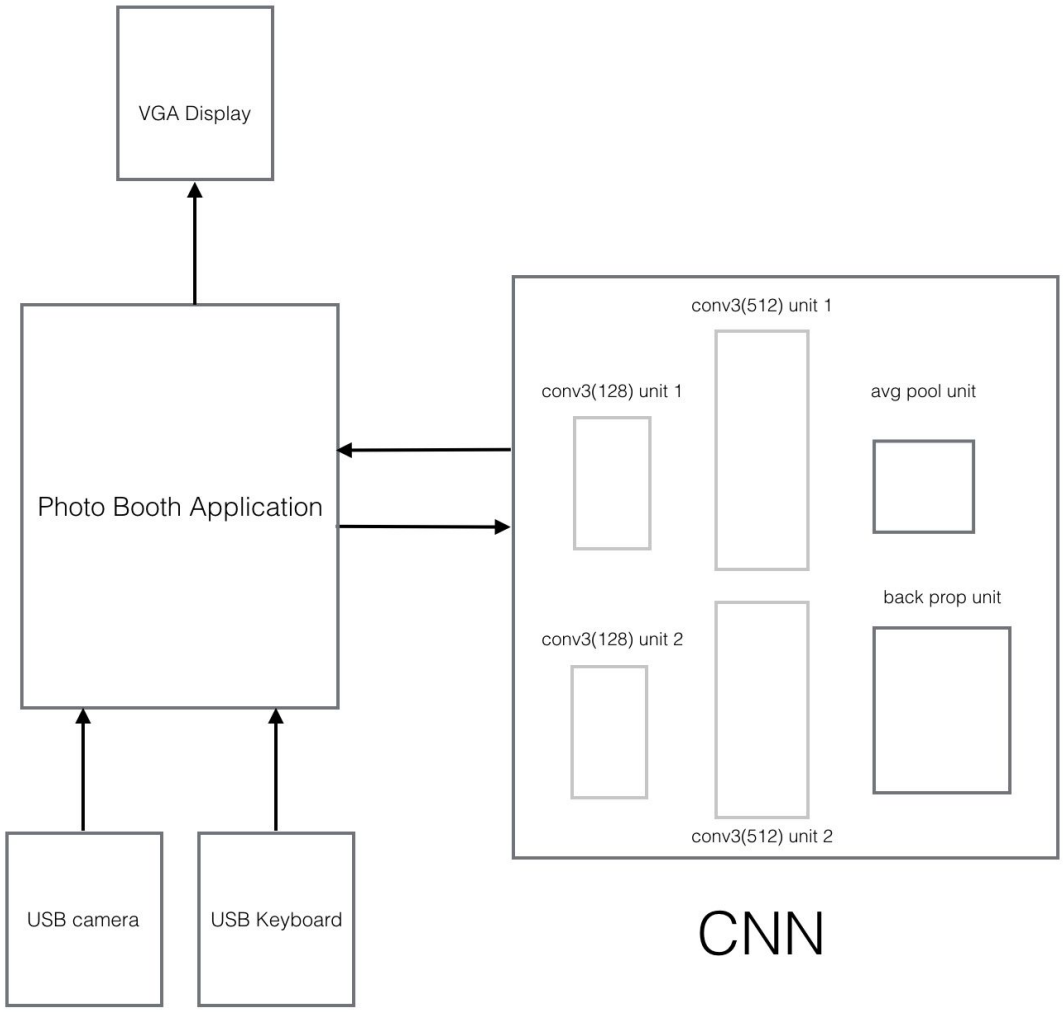


1

With FPGram, users will be able to produce personal images as though they have been created by their favorite artists.

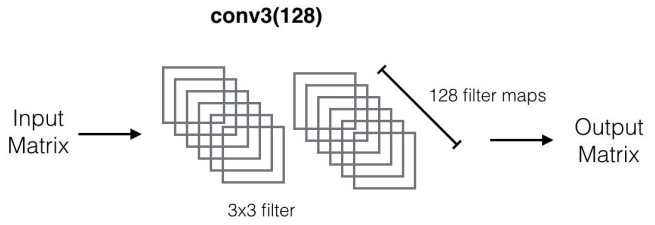
1. Gatys, L.A., Ecker, A.S. & Bethge, M. *A neural algorithm of artistic style.* http://arxiv.org/abs/1508.06576 (2015).

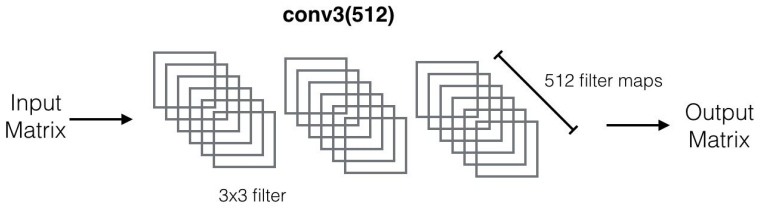
Design Overview



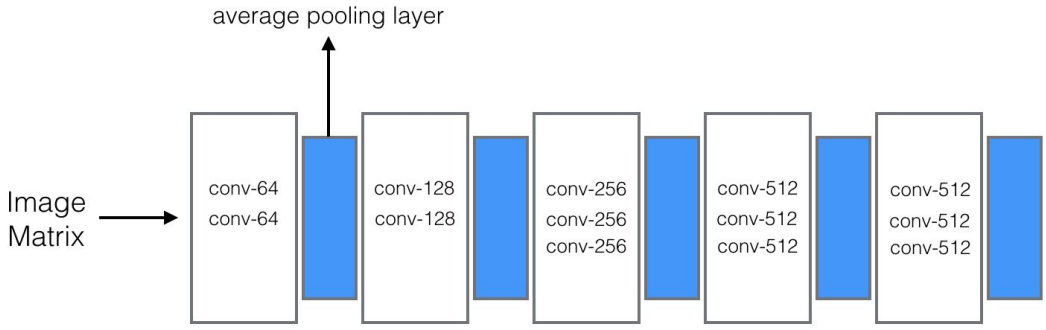
Convolutional Neural Network on the FPGA

The CNN device will mostly be implemented in hardware. It consists of 4 units of stacked convolutional layers, an average pooling unit and a back propagation unit. Each of the convolutional layer units use 3 sets of 3x3 filters with units have 128 output feature maps and the others have 512 output channels.





The images go through 5 rounds of convolution. The first round uses 2 sets of 64 filters from the conv3(128) unit. The second set uses the full bank of filters in the conv3(128) unit. The third round uses 3 sets of 256 filters from the conv3(512) unit and the remaining two sets make use of the full bank of the conv3(512) unit.



VGG Convolutional Neural Network

After each round of convolution, the feature map is passed to the average pooling unit r to downsample the image. One set of convolutional layer units are used for the input image and the other set on a white noise image that will generate the desired output. The standard error of the back propagation of the image is used as the gradient for the image to generated. The back propagation unit performs two tasks. It finds the gradient for both content recognition and style recognition. The CNN device should already be trained for style representation. We are able to derive our desired output by jointly minimizing the distance of the white noise image from both the style and content representations.

Software: Photo Booth Application

The application will consist of a simple interface which allows a user to either upload a picture or take a picture with the USB camera, select an artistic style from a bank of several popular artists and a button to begin processing the new image. Once the image is done processing it will be displayed on the VGA display.

Device Drivers: USB Camera, USB Keyboard, VGA display and interfacing with the CNN

The software will be interfacing with four different devices. The USB camera and VGA display act as inputs to the program. The read and write functions receive and send a matrix of pixels of 3 channels representing the RGB values. This will follow the avalon bus protocol. The framebuffer for the VGA display will be written with software.

Timeline

Milestone 1 : Hardware Implementation

- Implementation of convolutional layer units
- Implementation of average pooling unit
- Implementation of backpropagation algorithm

Deliverable: SystemVerilog files for each unit

Milestone 2: Training

- Connect convolutional neural network layers and start testing
- Write device driver and implement bus protocol for CNN
- Train the network for preset style representations

Deliverable: Demo of neural network device

Milestone 3: Complete System

- Write drivers for keyboard, camera and VGA display
- Finish software application for user interface

Deliverable: Demo of system in addition to software and hardware files

References

1. Gatys, Leon A.; Ecker, Alexander S.; Bethge, Matthias. *A Neural Algorithm of Artistic Style.* arXiv:1508.06576 [cs.CV].

2. Karen Simonyan, Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition* arXiv:1409.1556 [cs.CV].