Automated Replication of Art Styles in Shaders through Generative Adversarial Networks

Student name: Curtis Wiseman

Student ID number: 15012838

Supervisor name: Tom Bashford-Rogers

Date submitted: 26th of April 2017

Module: UFCFS4-30-3 Creative Technologies Project

Screenshot/ image of the work (600 pixels high x 800 pixels wide .jpg @ 240 dpi)

Summary

A combination of the deep convolutional network technology of a generative adversarial network with the linear optimisation capabilities of a genetic algorithm to manipulate pre-set shaders into approximating a given cartoon-like art style. The logical systems are written in Python 3.6 and the rendering is performed in Unity 2017.1.

Brief biography

Since a young age I have always appreciated the visual aspect of games and films, often being just as inquisitive about the technical achievements in this regard as the artistic ones. I intend to pursue technical art for 3D animated media in my own career and being able to offer a tool that provides a layer of mutual communication between the creative artist and the technical programmer would represent my strengths as a bridge between specialists.

How to access the project

Please put down the URL and/or details of how to access your project, and the ***URL of your final video***. If we need to see any back end / administration interface, please provide us with necessary passwords and URLs to access this. Basically, what we cannot access will not be considered for marking. You can change the passwords after you receive your mark.

We also want to be able to see source code, and the best way is to download it from your site or GitHub. Please clearly comment code to show us what is your own and what has been used from frameworks, libraries, OSS or borrowed from elsewhere.

If there is some other method for providing access to your project you will need to provide instructions here. Also, if there is anything we need to know about the work that will not be self-explanatory, then also provide brief instructions here.

Introduction

Many 3D video games and animated films try to replicate the aesthetics of 2D animation. Films may simply be doing so as an artistic choice, like with the [year] film [film]. This can also happen in video games, as in Clover Studios’ Okami, but oftentimes games are trying to mimic a cartoon or comic book that the game is based on. Examples of this include [example] and [example].

The issue therein is that 3D animation is fundamentally different to 2D animation in many ways, and so replicating the style of one in the other can be difficult. 3D animation may have been popularised in the 1990s by the likes of Pixar, but even to this day studios struggle to bridge the gap between the two mediums. To do so, shaders must be created to mimic 2D hallmarks such as limited shading, linework, and hatched shadows. The most impressive attempts, such as those shown by Arc Systems Works in their fighting games Guilty Gear Xrd and Dragon Ball FighterZ, take so much time and effort to craft that it may even be more economic to authentically 2D animate them.

The purpose of this project then was to create a tool that could be used by these studios to automate much of the shader creation process, freeing up the studio’s workforce to focus on other tasks. It should be able to take in a set of images displaying the 2D art style wanting to be replicated, and, within a reasonable timeframe, produce a set of shaders. These shaders, when used together in a 3D rendering program, will give the impression that the 3D objects being rendered are in fact 2D drawings.

It was also an exercise in experimentation, to see if it is truly possible for a machine to learn to replicate art styles more efficiently than a human. The recent upswing in machine learning research shows that the focus has shifted, from building smarter artificial intelligence (AI) machines to building ones that can make themselves smarter. It is possible that in the coming years, AI will have a much larger hand in translating artistic intent into a digitally replicable form than ever before.

All in all, the objective of this project was to produce a relatively easy-to-use program that could be set running and left, with the user returning later to find a convergence on an appropriate shader configuration. A Python program and Unity application are to be used in conjunction for the full pipeline to be functional.

Practice

State the main outcomes of your practice. Insert graphs, screenshots, diagrams, short code snippets if necessary.

A program has been produced that adequately performs the task set for it. Tested on both frame data extracted from episodes of [TV show] and an 8000-page archive of the comic *Homestuck*, both known for their distinct art styles, it configures shaders that mimic them a closely as possible within the parameters of the shaders provided. Given robust enough shader parameters and plentiful training time, the configurations produced would theoretically converge towards accurate replication. The data shown by the limited training performed does show a trend towards better mimicry of the styles. The biggest obstacles preventing further progress are the base shaders being used. They are simple and do not incorporate advanced effects such as normal manipulation or post-processing effects like bloom lighting or colour grading. Expanding the repertoire of available effects would greatly enhance the program’s ability to produce more useful configurations. This would require a large amount of human input, but the base shaders would only need to be written once, thus meaning that less human intervention would be required once a significant shader library is built up.

Initially, it was intended for the program to configure two ubershaders; one would be applied to dynamic or “foreground” objects and the other to static or “background” objects. This idea was dropped for the sake of time, and because of a general move away from ubershaders. The reason for the move to instead use a collection of regular shaders was to increase compartmentalisation and customisability by allowing for the end user to outright disable or enable certain effects as they please. If an effect in a ubershader is not being used, then either it is still being calculated and multiplied by zero or there is a conditional branch that averts the code section entirely. Both are relatively costly procedures when compared with simply removing a specialised effect shader entirely. However, the program can still be used with an ubershader if the user has one prepared. Aside from the shader itself, the only other input needed is an XML file containing the names of all the shader properties to be trained on by the program. The properties must be represented as floating-point numbers normalised to the range 0-1, but within the shader itself this can be multiplied to a more fitting range if needed.

The main program was written in Python 3.6, which was chosen because of the useful machine learning libraries that have been made for it. One used for this project is TensorFlow, the tool of choice for producing generative adversarial networks, also known as GANs. It was decided to use a GAN because they are an emerging technology and a hot topic for machine learning computer scientists at the moment. GANs are networks that generate new data mimicking the controlled input data by pitting two competing networks against each other, one learning to produce more convincing data while the other learns to discern the difference between real and generated data. The capabilities on GAN learning are being pushed on many fronts, but during the research phase, no evidence of them being used to generate shaders was found. This opportunity to explore a relatively uncharted field of research was a major motivational factor in pursuing it.

However, GANs are defined in such a way that they specifically generate data in the same format as the input data; if the input data is an image set, the generated data will all likewise be images. This is a problem when trying to produce shaders, as their effects are visual and therefore best represented by images, while the data that controls them is parameterised and made of floating-point number variables. These two conflicting formats would not work together in a normal GAN, as became clear during my research. Instead, it was decided to modify the underlying structure of the GAN to suit the two formats. The discriminator network, that classifies data as real or generated, is already suitable as shaders are judged on the appearance of objects rendered with them. The generator network required a complete overhaul though. Most of its functionality was replaced with a genetic algorithm. Genetic algorithms (GAs) are loosely based on the theory of evolution, wherein solutions, represented by individuals in a larger “population” of solutions, are at first randomly generated. They then are tested by some metric to decide their “fitness” and the most fit individuals are cross-referenced to produce “offspring” individuals. Some randomly-picked individuals are selected for this “breeding” process as well for the sake of diversity and to avoid reaching a sub-optimal solution, and parts of individuals are sometimes randomly changed for this purpose as well. This implementation is a good fit for shader optimisation. The solutions are dictionary objects containing the names of different effects and a float value for each of them. The discriminator network’s judgement on how likely the solution’s rendered output is to be real makes for a well-defined fitness function. There is also a theoretical optimal solution, and genetic algorithms are guaranteed to find optimal solutions if their fitness functions are smart enough and if they have enough time to iterate through a sufficient number of generations.

There were, naturally, difficulties in combining the two approaches. A GAN implemented in TensorFlow deals almost entirely in tensors, an abstract data type that can be cumbersome to work with outside of functions specifically designed to accept them. The biggest hurdle in this regard was translating the output from the discriminator network into something the GA could use for its fitness value. For all its calculations, the GA requires that the fitness of its population’s individuals be a float or double, or similar high-precision basic type. Tensors are somewhat like multi-dimensional arrays. However, they do not make their values easily obtainable as they are not stored in the same way. Tensors act more like instruction sets, being evaluated on-demand within a TensorFlow session. This challenge was overcome by reducing the tensor down to zero dimensions, evaluating the value in a new TensorFlow session into a more easily-readable array and then copying the useful data into a new variable.

Another issue was rendering test images with the shaders being configured. It was decided early on that the rendering would be done in Unity due to its high compatibility with various shaders and ease of use. Unfortunately, Python code does not run natively in Unity’s C#-based pipeline. Because of this, making the Python code and Unity renderer work together presented a challenge. Both programs had to bridge the gap in different ways. The Python code would serialise the shader data into an XML file with a unique identification number as its first node. Unity, then, would be constantly monitoring for changes in that ID number, and reading the full file whenever a change was made. The file data would then be unpacked into the corresponding shaders’ run-time attributes. Because shaders can be updated in real time, Unity would then take a screenshot using the built-in screen capture method and save it to a new file with the aforementioned ID number as its file name. The Python program, at this point, would be anticipating a new file with that ID number to be created, and would therefore be ready to read and decode to be passed back to the discriminator network for fitness testing. There were issues wherein the Unity side would only update once, and the only workaround outside of restarting the program would be to click on a different window and back onto the Unity one every time an update was needed. This was eventually resolved by [please say I fixed this]

The three shaders used for testing were an outline shader, cel shading, and cross-hatching. These were downloaded from the internet but modified to be easily configurable via trainable float values. The reason these three were chosen is because they are visually distinct but quite common in drawings and animation. Outlines and cell shading are more common in dynamic objects as they are easier to draw many times, whereas cross-hatching is rarely used on characters and is more often found in background art. However, it is possible to use all three together, each with different weights and colours. While a broader range of effects would have made the system more versatile, that is something more easily implemented by the end user. The system is left open so that new shaders can be added by adding them to the Unity project and providing entries for them and their trainable values in the XML file. After all, new shader techniques are being developed all the time, so leaving the door open to future possibilities is a must.

Discussion of outcomes 900

Reflection: What significance does your completed project have in a wider professional context?  
What can other professionals draw from your work?  
What does it improve on?  
Critically analyse what you have achieved and if necessary suggest different future approaches.   
To what extent do the outcomes of your practice satisfy your original objectives, stated in your proposal?  
e.g. have you discovered any new methods as you went along?  
New models that might help others understand processes better?  
Developed a good protocol for implementing things?  
Streamlined some processes?

The research done here proves that GANs are not limited to image generation, nor do they need be constrained by having inputs and outputs of the same format. While it may not mean much in the wider scope of machine learning, small steps like this help ensure that progress is being chipped away at. This kind of research may open the door to future endeavours for machines to learn processes rather than outcomes. A machine could learn to replicate a style of music from the instruments up, or to design house floor plans in an aesthetically-pleasing manner by placing furniture and appliances individually. Having building blocks, human-designed but modifiable elements at its disposal, the machine would have a much more accurate set of tools to work with when compared to traditional neural network generation. In a regular neural network, the machine tries to guess the construction of an object from the pixels, or other primitives. This leads to a lot of networks failing to recognise spatial cohesion, whereas a network that builds solutions from more easily-labelled parts would have a better understanding of how elements interact to form a whole. The total number of possible outcomes would be reduced, but the ones remaining would be more likely to pass a human-run quality test.

With that said, the implementation here is not quite so impressive. Its outputs are extremely limited due to the small number and simplicity of shaders used. It has the potential to be a broader and more robust program given further work, but it is more substantial in theory than in practice. Combining this tool with advanced shaders such as the ones found in the SIGGRAPH ’03 paper Coherent Stylized Silhouettes could lead to extremely style-accurate shaders with minimal human interaction.

The project has changed course since the original proposal. Initially, it was intended to simulate the motion of 2D animation in 3D by using reverse kinematics to alter animations in real-time. This proved to be too ambitious for the scope of this project. As explained by Arc Systems Works technical artist Junya Christopher Motomura in a GDC talk, simulating the motions of imprecise brush strokes is best done manually, and even then is a highly time-intensive task. The decision to take the project in a different, more machine-learning-orientated direction greatly increased its feasibility.

Conclusion and recommendations

Overall, the project is a theoretical and research success, but the practical implementation is barebones and lacking in substance. More work needed to be put into polishing the user experience and better showing off the program’s capabilities.

The possibilities of combining machine learning types open up many more opportunities for developers to find hybrid solutions – much like the breeding function in a genetic algorithm, the best qualities of different approaches could combine to make one greater than the sum of its parts. Perhaps, in the future, it may even reach the point where a machine learns how to produce better machine learning algorithms itself.

This project, in particular, has the potential to be a powerful tool if given proper care and attention past its initial release. With a large library of shaders representing a breadth of effects at its disposal, it could be used to automate much of the process of turning an artist’s vision into dynamically-rendered visuals on the screen. Artists and programmers alike could use it to bridge the gap between their specialities.

References (not included in word count)

Author, A. (2009) *A Book About Student Projects*. Location Publisher.

Author, B (2008) ‘Journal Article’, *Digital Media Journal*, Vol 1/13, pp 13-23

University of the West of England (2009) *UWE Library Services: Study skills - The Harvard System* [Online] Available from <http://www.uwe.ac.uk/library/resources/general/iskillzone/referencing/harvardreferencing/> [18 September 2009]

Bibliography (not included in word count)

Appendixes: (not included in word count)

A Log sheets

B Project timeline

C If necessary, insert a list of interview questions, any larger tables, evidence of design development, longer code snippets or other relevant materials in here. This does not come under the wordcount. Nevertheless, only insert useful materials here, please don’t bulk it up. Your report should be able to stand on its own, without relying on appendixes.