Robert Chang, Frank Xing

Aditya Joglekar

Section 7 35L

28 November 2019

Deep Learning In Video Games

Traditionally, animations in video games were done mostly by hand, which is a logical and effective method when paired with 2-D cartoonish games. However, with the recent influx of 3-D games, the flaws of this approach became apparent, like unrealistic and incomplete details regarding surface dynamic of materials such as hair and cloth. Eventually, there was a further breakthrough with motion capture technology and it being able to capture small cinematic cutscenes with accuracy. Unfortunately, this still can't account for every possible interaction between characters and environments. Overall, these techniques discussed above are limited and not ideal due to also being too expensive and time-consuming.

The article *Deep Learning Is Making Video Game Characters Move Like Real People* introduces a group comprised of computer scientists from the University of Edinburgh and Adobe Research that devised a possible solution. Their solution utilizes a Neural State Machine that takes in large databases of various performance capture animations and subsequently uses a deep neural network in accordance with numerous data driven techniques to create smooth transition animations between different states. This is incredibly useful because instead of having to create a separate animation for every single possibility that may arise in a game which may amount to tens of thousands of animations, their solution can automatically, based on the amount of inputs received, fill in the gaps between different positions of characters to create its own

animations, thus saving time and resources. This solution also has a modular aspect to it that grants its user even greater flexibility. It uses a deep neural network which contains multiple hidden layers and numerous intermediate mathematical equations that map such relations between layers. When databases are inputted, multiple possible outputs with different weights can be generated and the user can actually end up with even more transition animations outputted than he/she initially was looking for. This further enhances the solution's practicality since project developers can obtain excess animations and could therefore save even more money and resources. The researches and computer scientists behind the solution effectively describe their creation by emphasizing two features: planning and adaptation. The planning feature involves the network figuring out and planning what kinds of different movements are needed to create the desired action. The adaptation feature consists of the solution's ability to adapt and adjust to changes in the game environment and to correct the animations being generated, adding to its already extensive flexibility.

Our initial reactions mainly consisted of feelings of curiosity, interest, and inspiration. We both felt that the article and subject matter were both incredibly interesting because we could relate to it on a personal level since we both grew up with video games.. It was an eye opening article since we never actually realized the possible uses of deep learning in the video game industry in conjunction with animation. Looking back, we felt that it was an incredible testament to how fast and far technology has developed in the past decade. Who could have imagined basic pixelated games on our Gameboy systems to become games that use deep neural networks to automate certain stages of the animation process - a feat achieved in merely two decades. However, although the paper was incredibly inspiring and gave us a lot of insight into the future

of such technology and its implications for the future of the game industry, we felt that the paper lacked technical details since a lot of the explanations for how the solution actually worked was unmentioned. The explanations were more concept oriented rather than technologically detailed and so we do not have a full understanding of the specific mechanisms. However, it is important to note that the reason behind is that they haven't actually formally presented their research to the world yet. They will be presenting it at the end of this year at the ACM Siggraph Asia Conference and so we expect more details to be explained then. Although it lacked specifics, we still thoroughly enjoyed and appreciated the article since it increased both our interests in technological applications in the entertainment industry and we will both definitely explore that field in the future.

With our interest in neural networks and deep learning piqued, we used our online resources to search for related works that utilized other technologies or approaches in order to solve similar problems. We found the Head Motion Synthesis project which used deep neural networks (DNNs) that depend on audio-visual inputs to predict head movement of a speaker based on his/her speech. The prediction of a single speaker's movements were improved if there were other speakers on the stage and the results could be used in talking avatar cutscenes. Like the solution presented in our article, this project's solution also takes in input and subsequently outputs an automatically generated animation. We also found the Generalized Speech Animation, which targets mouth animation in order to realistically create human motion movement for animated speech, replying on a DNN that takes audio-visual input to achieve its results. Additionally, the technique also requires minimal tuning, generalizes well to speech pattern and can be edited for emotional effect. There's also the Facial Performance Capture

which relies on visual input captured by motion capture technology in order to feed a DNN that uses an algorithm to generate and fill in the blanks for refined facial 3d recreation with the geometry and textures of the environment. The last related work that we found that shared similar technological approaches or techniques was the Action-Conditional Video Prediction project which also used a DNN. Instead of inputting just any motion capture video, it can take in any image frame and transforms it in correlation with various variables and actions to generate a future image. This is incredibly beneficial because this presents the user with the ability to generate visually-realistic frames that can be used again for the next 100 or so frames. This project's solution is extremely flexible and can be applied to any sort of visual input in order to generate some sort of future frame based on the client's/user's demands. The fact that all these different applications share the same core technological principles just goes to show the wide extensibility of deep learning and its many implications for the future of technology.

This article also led us to think about even more possible applications for such technology in the future. We thought that with the next generation of gaming, which is now online game streaming platforms that directly allow users to stream game data to their computers from online servers, can be greatly improved with this new technology. The main reason we believe this to be the best application is because since this solution reduces the number of animation clips that are needed which in turn lowers the amount of data that needs to be processed, making streaming become a more viable option. This deep learning solution can also increase future video game world sizes since a lot of animations can be streamlined to neural network processes rather than animated beforehand. Other applications involve different entertainment industry uses such as CGI for movies, animated shorts, and tv shows. This

technology is also applicable to improve the navigation and camera systems of self-driving cars or any software that relies on the construction of visuals from a set of patterns or other graphics. Visual and augmented reality are also possible applications in addition to environment analysis. Away from the entertainment industry, the military can also greatly benefit from such a solution since this can greatly automate and aid the generation of tactical plans, strategies for military operations, movement patterns, entry plans, analysis of enemy movement routes, and infiltration simulation software.

There are simply too many possible applications for such a finding and we both believe that it will be a game changer for the entertainment industry. We both drew inspiration from this article and now believe that artificial intelligence and deep learning have just begun to take flight and within a couple of years, we will be staring at a new and unimaginably advanced world, where adaptation and improvements occur at a rapid pace.

Works Cited

"A Deep Learning Approach for Generalized Speech Animation." ACM Transactions on Graphics (TOG), ACM, July 2017, https://dl.acm.org/citation.cfm?id=3073699, https://doi.org/10.1145/3072959.3073699.

Codrina. "Skeletal versus Sprite Sheet Animations: Advantages and Disadvantages ." Marionette Studio, 20 Oct. 2016, https://marionettestudio.com/advantages-and-disadvantages-of-skeletal-versus-sprite-sheet-animations/.

Ding, Chuang, et al. "Head Motion Synthesis from Speech Using Deep Neural Networks." SpringerLink, Springer US, 24 July 2014, https://link.springer.com/article/10.1007/s11042-014-2156-2, https://doi.org/10.1007/s11042-014-2156-2.

Guo, Xiaoxiao, et al. "Action-Conditional Video Prediction Using Deep Networks in Atari Games." Action-Conditional Video Prediction Using Deep Networks in Atari Games, 1 Jan. 2015, https://papers.nips.cc/paper/5859-action-conditional-video-prediction-using-deep-networks-in-atari-games.pdf.

Liszewski, Andrew. "Deep Learning Is Making Video Game Characters Move Like Real People." Gizmodo Australia, Gizmodo Australia, 31 Oct. 2019, https://www.gizmodo.com.au/2019/10/deep-learning-is-making-video-game-characters-move-like-real-people/.

"Realtime Performance-Based Facial Animation." ACM Transactions on Graphics (TOG), ACM, July 2011, https://dl.acm.org/citation.cfm?id=1964972, https://doi.org/10.1145/1964921.1964972.

Starck, Jonathan, and Adrian Hilton. "Surface Capture for Performance-Based Animation." Surface Capture for Performance-Based Animation - IEEE Journals & Magazine, 7 May 2007, https://ieeexplore.ieee.org/abstract/document/4178157 , https://doi.org/10.1109/mcg.2007.68.

Theofficialacm. "Deep Neural Network Generates Realistic Character-Scene Interactions." EurekAlert!, 29 Oct. 2019, https://www.eurekalert.org/pub_releases/2019-10/afcm-dnn102919.php.