Legend of Wrong Mountain: Full Generation of Traditional Chinese Opera Using Multiple Machine Learning Algorithms

Lingdong Huang Carnegie Mellon University lingdonh@andrew.cmu.edu Zheng Jiang Carnegie Mellon University zjiang1@andrew.cmu.edu Syuan-Cheng Sun Carnegie Mellon University syuanchs@andrew.cmu.edu

Tong Bai Carnegie Mellon University crystalbai1994@gmail.com Eunsu Kang Carnegie Mellon University kangeunsu@gmail.com Barnabas Poczos Carnegie Mellon University bapoczos@andrew.cmu.edu

1 Introduction

As one of the oldest forms of Chinese Opera since the 16th century, *Kunqu* features literary virtuosity in its scripts, sophisticated vocal techniques in its singing, emotional and elegant yet rigorous bodily motions and facial expressions in its performance. Over its history, *Kunqu* has developed established modes and patterns, which makes it especially suitable for neural networks to learn. In order to generate *Kunqu* literature and performance computationally, we applied machine learning techniques such as LSTM [4], pix2pix [5], pix2pixHD [9], RNNs [2], Markov Chain, OpenPose [7] and Detectron [3], as well as tools such as Canny Edge [1], PIL and SketchUp. Our result, Legend of Wrong Mountain, is novel twofold: it is the first and only generated Opera at the time of this writing, and it is a machine's attempt at *Gesamtkunstwerk* [8], "the Total Artwork". It explores the marriage between contemporary technologies and traditional art form. By studying historical scripts, musical notations and traditional methods for creating *Kunqu*, we tweaked existing algorithms and devised new ones to conform to the traditional rules and norms as closely as possible. We present this project as a video accompanied by audio.

2 Methods

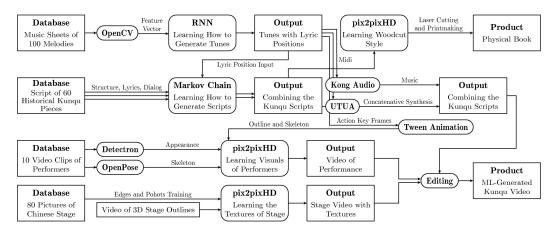


Figure 1: Workflow Pipeline

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

2.1 Music Generation

The music of *Kunqu* are composed of melodies with established patterns. A hundred images containing traditional *Kunqu* sheet music were downloaded from *Kunquwang* (http://www.52Kunqu.com) using a scraper script. The sheets use Numbered Notation (Ziffersystem), a common format for notating Traditional Chinese music, which we wrote a new computer-vision based algorithm to parse. Then we used charRNN([6]) to generate our unique tunes. Instead of using the conventional method of encoding note-down-per-tick as a 128D vector, we concisely represented them with descriptive ASCII strings. For example, middle C is C_4. The pitch class information makes the structure easier for the learning model to grasp.

2.2 Script Generation

To learn the hierarchical structure of Kunqu scripts, we introduced a nested system with different Markov Chains at each level. The method is capable of producing outputs containing reasonable large-scale structure for chapters, dialogues, prompts and lyrics, as well as smooth sentences within the structure. For example, the sung parts of Kunqu were typically written in Qu, a form of classical Chinese poetry consisting of characters that conforms to tonal rules from melodies. Firstly, we reversed the chain for each sentence, such that the last, rhyming character is first determined, and all the previous characters are predicted backwards. Secondly, we inform the Markov Chain with tonal rules inferred from the generated melody. Thus, the generated poems not only were read smoothly but also rhyme and adhere to the traditional tonal rules.

2.3 Visuals Generation

To generate graphics for the performers, we first used PIL to render new poses from scratch by implementing a simple pesudo-3D engine that contains skeleton system, physics and cloth simulation. Next, the procedurally generated poses were fed back into pix2pixHD to re-generate the photorealistic visuals of the performers. As we had 9 pix2pixHD networks, each trained on a differently costumed character from the input videos, the network output 9 animated photographic performers that corresponds to the procedural skeletons and outlines. On the other hand, to generate the scene, We built a 3D model of the stage using SketchUp and wrote a Ruby script to control the camera. Simple outline drawings were exported as image sequences and sent to pix2PixHD. The neural network is trained using 80 photographs of historical opera houses against their extracted outlines produced by Canny edge. As the output, pix2pixHD automatically textured our line drawings with the stylistic elements of the training set. At last, the generated images were combined into a video tour of the opera house at the start of the performance.

3 Conclusion

This paper presents our techniques in generating Chinese Opera using machine learning methods. The project was presented at the Art + Machine Learning exhibition in Pittsburgh on May 1st, 2018 and received overwhelmingly positive feedback. People were fascinated by the mesmerizing performance as well as the curious story. The characters in the opera script were rendered in the style of traditional Chinese woodcut using pix2pix, and printed using historical woodblock press techniques on hand-bound books, which were offered to the audience at the show as further enhancement of the experience.



Figure 2: Renderings of Stage and Actors

References

- J Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, June 1986.
- [2] Felix A Gers and Jürgen Schmidhuber. Recurrent nets that time and count. In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, volume 3, pages 189–194. IEEE, 2000.
- [3] Ross Girshick, Ilija Radosavovic, Georgia Gkioxari, Piotr Dollár, and Kaiming He. Detectron. https://github.com/facebookresearch/detectron, 2018.
- [4] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [5] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *arxiv*, 2016.
- [6] Andrej Karpathy. Char-rnn, 2015.
- [7] Vahid Kazemi and Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1867–1874, 2014.
- [8] Karl Friedrich Eusebius Trahndorff. Ästhetik oder lehre von weltanschauung und kunst. 1827.
- [9] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. arXiv preprint arXiv:1711.11585, 2017.