Animation Video Resequencing with a Convolutional AutoEncoder

Shang-Wei Zhang
National Cheng Kung University

Charles C.Morace
National Cheng Kung University

Thi Ngoc Hanh Le
National Cheng Kung University

Chih-Kuo Yeh
National Cheng Kung University

Sheng-Yi Yao
National Cheng Kung University

Shih-Syun Lin
National Taiwan Ocean University

Tong-Yee Lee
National Cheng Kung University

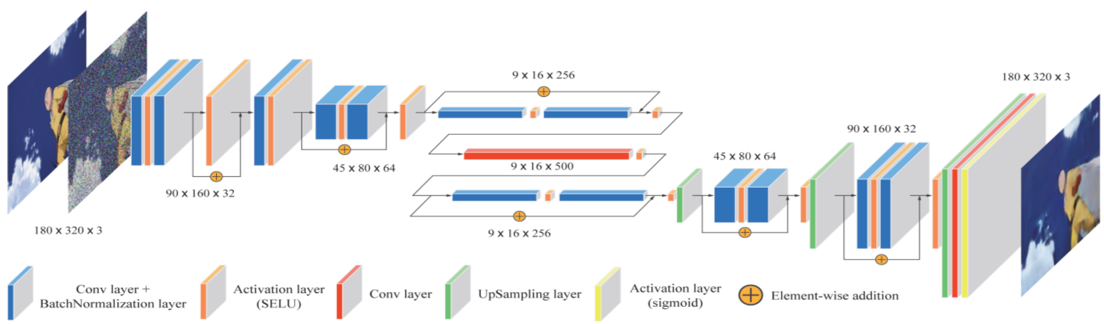

Figure 1: The network architecture of the proposed autoencoder.

CCS CONCEPTS

• **Computer methodologies → Animation**

KEYWORDS

Nonlinear Dimension Reduction, Manifold learning, Animation Sequencing and Resequencing

ACM Reference format:

Shang-Wei Zhang, Charles C.Morace, Thi Ngoc Hanh Le, Chih-Kuo Yeh, Sheng-Yi Yao, Shih-Syun Lin and Tong-Yee Lee. 2019. Animation Video Resequencing with a Convolutional AutoEncoder. In *Proceedings of SA '19: SIGGRAPH Asia 2019 (SA '19 Posters), November 17-20, 2019, Brisbane, QLD, Australia.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3355056.3364550

1 INTRODUCTION

Animation has brought to life the creative potential of the human mind. To create more realistic looking animations, animators commonly utilize a set of principles, including natural movement, as a model and will incorporate other principles for dramatic effects and emotional impact. There have been many techniques developed to ease the computer animation pipeline, production is still an arduous process that involves the creation of many image sequences depicting the motion of complex characters and their environments. If a single image is out of place, the whole animation may be ruined by an unnatural movement, which is not only visually displeasing but also distracts from the narrative.

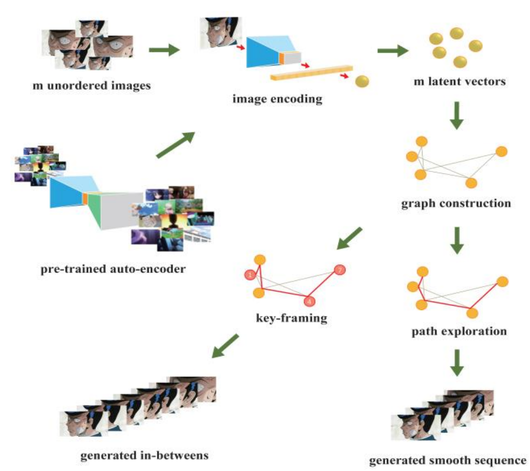

Figure 2: The system overview

SA '19 Posters, November 17-20, 2019, Brisbane, QLD, Australia

ACM ISBN 978-1-4503-6943-5/19/11.
https://doi.org/10.1145/3355056.3364550

Smooth images sequencing is the key to produce visually pleasing animation. Most of the previous researches divide the sequencing problem into two distinct steps: (1) establishing a suitable distance measure for the similarity between the input images; (2) determining an optimal sequence according to the similarity measure defined in the first step (Schödl and Irfan, 2000, Ling and Jacobs, 2007, Osadchy et al., 2007, Holden et al., 2015, Fried et al., 2017). In our proposed method, we use a convolutional autoencoder for both feature extraction and dimension reduction. In other words, our research learns a topological manifold embedded in the space of images where smooth paths on the manifold represent visually plausible animations.

2 OUR APPROACH

Given an unordered collection of images as input, our proposed system can decide suitable in-between images for a set of key-frames, or synthesize new animation sequences which are locally "as smooth as possible". To control the output, a user has the option to supply either a set of key-frame images or a starting frame and sequence length; then the system will create an animation with the desired properties.

Fig. 1 demonstrates the network architecture of our proposed autoencoder. We firstly train a denoising autoencoder (Vincent et al., 2008) on a collection of animation images downloaded from the internet. After training the autoencoder, we apply the encoding network to a collection of images supplied by the user to obtain lower-dimensional latent vector representations. From these latent vectors, we compute a Euclidean minimal spanning tree (MST), and the proposed key-frame pathfinding uses the MST to generate in-between images. To synthesize new animations of arbitrary length, we compute a path-connected proximity graph and employ a Monte Carlo method to find a path that is as smooth as possible.

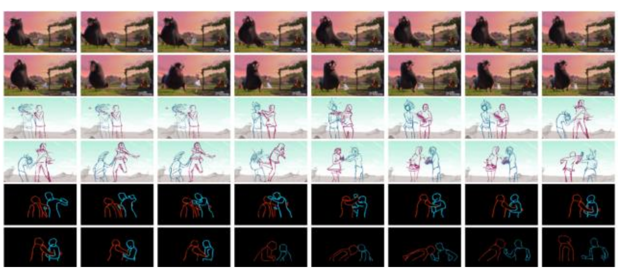

Figure 3: Results of our proposed key-frame method. The first and last frames are selected by a user and the in-between frames are generated by traversing the Euclidean minimum spanning tree.

To generate animation sequences, a key-framing method and a path exploration are proposed. In the key-frame pathfinding, the MST is employed to reduce in-betweens by traversing the path from one key-frame node to another. The paths connecting key-frame nodes in an MST are well suited for finding in-between images. In key-framing pathfinding, we return many temporally-coherent in-between frames to the user. And users may create animations from any number of key-frames by computing paths between consecutive key-frames and combining the results. In the proposed path exploration method, we employ a Monte Carlo technique for synthesizing new animations with the desired length.

3 RESULTS AND FUTURE WORK

In order to generate key-frame results, we examined the MST to guide key-frame selection and return precisely six in-between images. These results generated with the proposed method are shown in Fig. 3. Using the linear embedding of the MST for visualization provides a useful way to select key-frames that produce the desired number of in-betweens.

Fig. 4 shows the results produced with the proposed path-finding method. The path exploration method would generally produce better results for more extensive collections of input images containing a great variety of motion.

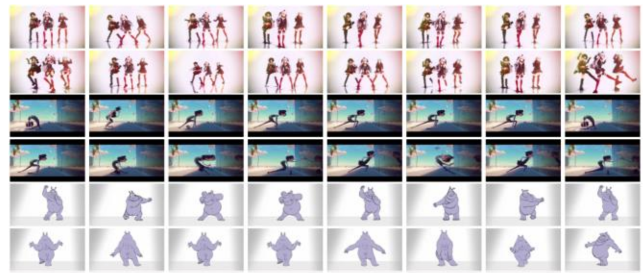

Figure 4: Results were created by randomly selecting a contiguous group of eight frames from an animation generated with the proposed path exploration method.

In the future, we would like to extend our work to a supervised learning framework which considers the temporal distance of the training data to solve above problems.

ACKNOWLEDGMENTS

We would like to thank the reviewers for many helpful comments. This work was supported in part by the Ministry of Science and Technology, Taiwan (107-2221-E-006 -196 -MY3, 108-2221-E-006 -038 -MY3, MOST-106-2221-E-019-069-MY2, and MOST-108-2221-E 019-038-MY2)

REFERENCES

Fried, O., S. Avidan and D. Cohen-Or (2017). Patch2vec: Globally consistent image patch representation. Computer Graphics Forum, Wiley Online Library.

Holden, D., J. Saito, T. Komura and T. Joyce (2015). Learning motion manifolds with convolutional autoencoders. SIGGRAPH Asia 2015 Technical Briefs, ACM.

Ling, H. and D. W. Jacobs (2007). "Shape Classification Using the Inner-Distance." IEEE Trans. Pattern Anal. Mach. Intell.**29** (2): 286-299.

Osadchy, M., Y. L. Cun and M. L. Miller (2007). "Synergistic Face Detection and Pose Estimation with Energy-Based Models." J. Mach. Learn. Res.**8**: 1197-1215.

Schödl, A. E., Irfan (2000). "Machine Learning for Video-Based Rendering." In Advances in Neural Information Processing Systems 13.

Vincent, P., H. Larochelle, Y. Bengio and P.-A. Manzagol (2008). Extracting and composing robust features with denoising autoencoders. Proceedings of the 25th international conference on Machine learning. Helsinki, Finland, ACM: 1096-1103.