Rotomation: AI Powered Rotoscoping at LAIKA

Jeff Stringer
LAIKA, LLC
Jstringer@laika.com

Narayan Sundararajan
Intel
narayan.sundararajan@intel.com

James Pina
LAIKA, LLC
jpina@laika.com

Swarnendu Kar
Intel
swarnendu.kar@intel.com

Nadine L. Dabby
Intel
nadine.dabby@intel.com

Adelle Lin
Intel
adelle.lin@intel.com

Luis Bermudez
Intel
luis.bermudez@intel.com

Sara Hilmarsdottir
Intel
madame.sara@gmail.com

ABSTRACT

In an ongoing collaboration, LAIKA and Intel are combining expertise in machine learning and filmmaking, to develop AI powered tools for accelerating digital paint and rotoscope tasks. LAIKA's stop-motion films, with unique character designs and 3d printed facial animation, provide challenging use cases for machine learning methodologies. Intel's team has focused on tools that fit seamlessly into the workflow and deliver powerful automation.

ACM Reference Format:

Jeff Stringer, Narayan Sundararajan, James Pina, Swarnendu Kar, Nadine L. Dabby, Adelle Lin, Luis Bermudez, and Sara Hilmarsdottir. 2020. Rotomation: AI Powered Rotoscoping at LAIKA. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Talks (SIGGRAPH '20 Talks), August 17, 2020.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3388767.3407315

1 INTRODUCTION

Stop-motion animation is a very precise and detailed artistic craft. Conjuring realistic performances from a puppet is painstaking, and traditionally limited, by mechanical armatures, rigid materials, and gravity. Since 2009's Coraline, LAIKA has been dedicated to removing these limitations and expanding the storytelling power of stop-motion while maintaining its dedication to the craft. Laika pioneered the use of 3d printed facial shapes, giving their puppets more natural and nuanced performances. While these innovations allow for magical in-camera animation, they also come at a cost for the visual effects team and particularly, the roto and paint team at the end of the pipeline. Missing Link pushed the use of 3D printed faces to a new level, creating bespoke performances for every shot in the film and resulting in more than 106,000 unique printed expressions. Each puppet's face requires retouching and frame by frame rotoscoping to remove seams and chatter.

Rotoscoping is still largely a manual process involving a suite of tools, and a significant investment of time by artists. Recent research in machine learning and image segmentation, has produced numerous papers, and a few products that promise to accelerate, or even automate the task of digital roto and paint. The collaboration between Laika's visual effects team and the Applied Machine Learning team at Intel is focused on developing systems and solutions that leverage the power of AI and ML techniques and seamlessly integrate into the artist's workflow.

The initial focus of our collaboration was on the removal of the seams between the mouth and brow shapes of the puppet faces. The task is consistent and repeatable, and relies upon key point detection of facial features. To use machine learning, a new, data efficient training model, specific to each character, had to be developed. Intel and Laika teams collaboratively designed a way to efficiently use data from as few as 5 completed shots along with renders, generated as a by-product of LAIKA's digital character design workflow, to create custom training models. A proof of concept for Missing Link's hero character, Lionel, was able to locate and label landmark tracking points and predict roto shapes for holdout mattes that could accelerate the roto work on most shots. The AI tools were embedded in a Nuke plugin that produced iterable results for the artist that improves as new shots were completed.

2 INITIAL INVESTIGATION AND SCOPE

For any AI enablement of existing workflows, it is critical that the solutions developed integrate seamlessly into existing pipelines with the ability for the artists to interact with the output and have final creative control. Towards this, the Intel and the Laika teams began by carefully reviewing before and after shots, along with the workfiles from the roto tasks to define the AI tool intervention. Specifically, the rotoscoping of temple and nose bridge seams for removal, as well as eyes and eyebrows for creating hold-out masks that can be pasted back into the frame after the seam is removed.

They also observed the roto and paint team's existing workflow and tools, identifying that the iteration on shapes and use of garbage mattes was key to the workflow. They did not expect that the output of any AI solution for key point tracking or shape prediction would be successful 100 percent of the time, so the toolset would need to preserve this level of artist interactivity to succeed.

3 A DATA EFFICIENT TRAINING MODEL

Generating these roto shapes through AI would typically require training on millions of face images with labeled key points, and a similar cache of "ground truth" images of completed roto, to help define successful holdout masks and roto curves. While pre-trained facial landmark detection models using human faces exist, those models do not generalize to puppet faces. Similarly, at the start of a production there are only a handful of completed test shots to represent the "ground truth" for a character. Thus, a major design

SIGGRAPH '20 Talks, August 17, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7971-7/20/08.
https://doi.org/10.1145/3388767.3407315

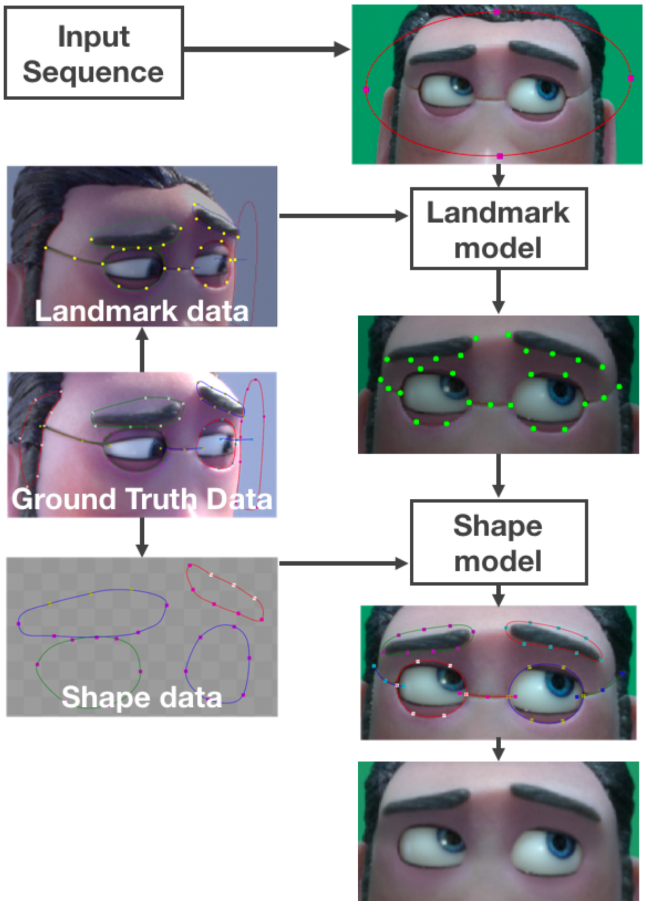

Figure 1: Auto-generation of tracking points and roto shapes

consideration for solving this problem is how to maximize the performance of a deep learning model in a data efficient way. The teams consciously explored techniques that harness what the studio is already producing, including pre-production assets that could be reused.

Pre-production at LAIKA includes designing the characters and establishing a range of poses in CG to be used in facial animation. The poses are printed and test performances are shot, and run through the entire roto process, for design approval. The character expression kits, created in Maya, can be rendered as photoreal images with labeled 3D tracking points on the geometry using some small modifications to our existing production pipeline. While the test shots generate a small batch of ground truth roto data.

Intel's team successfully developed ways to augment the small amount of ground truth data and CG renders of face kits, through random variations, to create the training data needed for a character specific neural network. With this efficient reuse of pre-production data, the neural network was able to reliably find the puppet faces in a shot and predict the position of landmark points for tracking the features for roto.

4 PREDICTING SHAPES FROM POINTS

The most challenging deliverable to automate is the prediction of useful shapes based on the landmark points detected. Shapes are predicted through end-to-end machine learning with RGB frames as input and finished rotoshape assets as ground truth output. An intermediate layer consisting of a few landmark points is introduced within the learning chain to enable an interactive workflow. The landmark points are handpicked to have the following two properties 1) visually consistent, i.e., can be located in terms of visible corners and edges in the RGB frame and 2) self-annotating, i.e., can be extracted based on certain geometrical and morphological operations on rotoshape assets, enabling inexpensive collection of ground truth data. Thereafter, two separate models are trained – a ResNet50-based network to predict landmarks based on a rough crop of the face region, and another custom Neural Network to predict roto shapes using only the inferred landmarks from the first model. The interactive methodology ensures that predicted landmark points can be adjusted for perfection before shape prediction is invoked.

5 WORKFLOW INTEGRATION

To be useful in a real production pipeline, these AI functions have to be wrapped up in a tool that fits into the artist workflow and gives them control over the results. Intel built a Nuke plugin with a simple, intuitive UI for the artist to try. The Intel tool is operated by the artist using only the source footage of the character and a user defined region. In minutes, rotoscope shapes and tracking data are delivered directly in the Nuke interface. The artist then has the option to use the data as is, modify it further in Nuke or export the shape and tracker data to SilhouetteFX. This allows the artist to iterate and generate more refined or new roto shapes and tracking data depending on the needs of the final paint composite integration. It also allows the artist to maximize the AI tool's potential directly within LAIKA's VFX production pipeline.

6 FUTURE WORK AND REFINEMENTS

In tests so far, LAIKA is seeing enormous potential for these tools to ease the heavy burden of post production in general, and roto and paint in particular, on their stop motion films. As the collaboration between Intel and LAIKA continues we expect to improve the accuracy of the initial results and improve the training models so that a higher percentage of shots can be solved for. We believe additional custom models can be trained to recognize full puppet bodies, common rig elements, and scanned props to further manipulate them.