Singed Silhouettes and Feed Forward Flames: Volumetric Neural Style Transfer for Expressive Fire Simulation

Paul Kanyuk
pkanyuk@pixar.com
Pixar Animation Studios

Vinicius C. Azevedo
vinicius.azevedo@disneyresearch.com
Disney Research Studios

Raphael Ortiz
raphael.ortiz@disneyresearch.com
Disney Research Studios

Jingwei Tang
jingwei.tang@disneyresearch.com
Disney Research Studios

Figure 1: A scene from *Elemental* where NST has been applied to the character and motorcycle fire. ©Disney/Pixar.

ABSTRACT

While controlling simulated gaseous volumes remains an ongoing battle when seeking realism in computer graphics, creating appealing characters entirely out of these simulations brought this challenge to an entirely new level in Pixar's film *Elemental*. For fire characters, like the protagonist "Ember", their faces and bodies needed to look and move like real fire, but not be so frenetic as to distract from the acting and emotion of their performances. Neural Style Transfer emerged as a key technique to achieving a look that met these criteria. By using more gaseous and languid pyro simulations as input, higher frequency "cusp and curve" shapes could be coherently transferred to the volumes via a GPU based optimization process applying recent advances in neural style transfer (NST) to the voxels themselves. These transferred shapes were controlled by hand painted styles and parameters that could be animated as a way to enhance the character performance. A key benefit of this process was the final shape of the fire was decoupled from the underlying simulation. This allowed the simulation to focus on lower frequency motions and stability, while NST could provide the final touches of shaping, particularly around the silhouette. Users could modify the perceived speed of the NST patterns by modulating the velocity input, and control where the effect was strongest by masking the resulting vector field. For large reusable simulations intended to be seen from many views, NST proved impractical as the multiple stylization viewpoints added excessive computational cost. To solve this, we trained a convolutional neural network to approximate the optimization process for stylization on representative volumes using multiple viewpoints. Then the much cheaper feed-forward neural network could be used, allowing us to bring NST to bear on larger scale environment simulations as well as characters.

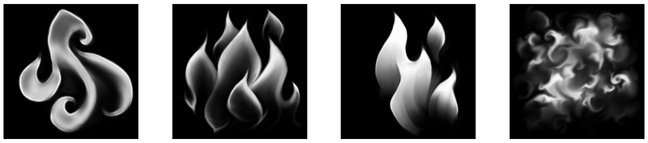
ACM Reference Format:
Paul Kanyuk, Vinicius C. Azevedo, Raphael Ortiz, and Jingwei Tang. 2023. Singed Silhouettes and Feed Forward Flames: Volumetric Neural Style Transfer for Expressive Fire Simulation. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Talks (SIGGRAPH '23 Talks), August 06–10, 2023, Los Angeles, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3587421.3595435

1 BACKGROUND

Early look development on Pixar's *Elemental* quickly ran into challenges when attempting to create characters made of fire. Flames at the scale of a large campfire were most visually appealing and recognizable as fiery, but were far too frenetic when filling the screen during a character performance. By contrast, fire at the scale of candle flame has a slow moving and mesmerizing quality that could support character action, but didn't have nearly enough visual interest and energy at the character's true scale. To find a solution, we tested 2d effects animation techniques and found that illustrated fire, while lacking realism, intuitively combined what was appealing from both campfire and candle flame. The illustrated fire moved with the lower frequencies of candles, but always emphasized "cusps and curves" in their silhouettes like a campfire.

SIGGRAPH '23 Talks, August 06–10, 2023, Los Angeles, CA, USA

ACM ISBN 979-8-4007-0143-6/23/08.
https://doi.org/10.1145/3587421.3595435



(a) Fleur (b) Plume (c) Ramps (d) Dissipate

Figure 2: Style images used for pyro NST. ©Pixar.

2 PARTICLES VS VOXELS

At the time of this exploration the FX team on the film *Luca* was working with Disney Research to test Lagrangian Neural Style Transfer (LNST) [Kim et al. 2020] for stylizing smoke simulations. Even though the technique was not ultimately used for the film, these experiments successfully demonstrated that illustrative styles could be transferred to physical simulations without sacrificing realism or coherence. Early tests with LNST indicated that a lower frequency pyro simulation could indeed be reshaped coherently to acquire the style of a target image. After extensive experimentation, a painted image of a swirling flame, dubbed the "fleur-de-lis" 2a, was our most successful style image for achieving the desired look, but there were a number of obstacles. By using particles as an intermediary for the Lagrangian approach, high quality results required extremely large point caches, and even then proved lossy. Speed was also an issue given the scale requirements. Our goal was to stylize the flames of the main character for every shot in active production within a 6 hour window in the late evening / early morning. Initial timing tests indicated each frame required 5 GPU minutes of computation on an NVIDIA P6000, which meant we only had one thirtieth the GPUs required.

When considering paths to optimization, we revisited LNST's voxel based predecessor, Transport-Based Neural Style Transfer (TNST) [Kim et al. 2019]. We discovered that the slowest computation of TNST, temporal coherence optimization, could be approximated with a far cheaper exponential moving average, reducing the computation time dramatically. This also eliminated expensive particle sampling and re-voxelization required for LNST. Coupled with the smaller voxel data footprint, the per frame computation actually dropped from 5 GPU minutes to 18 GPU seconds, a 16.667x speedup. This was combined with a hardware upgrade to NVIDIA A6000s, providing an additional 3x speed, totaling 50x.

3 THE ART OF TUNING NST FOR FIRE

While early tests with NST showed promise, parameter tuning was a journey. A main challenge with NST is finding the appropriate target styles to emphasize the desired features of the underlying input. Our favorite "fleur-de-lis" 2a style was very successful in accentuating vortices and swirls in the fire, but often created too many holes and circles. A second style image, the "plume" 2b, yielded excellent silhouettes and peaks, emphasizing buoyancy in the pyro, but missed the vortices and was a bit too vertical. Targeting both styles using two loss models for TNST successfully emphasized the desired features of each, but was actually too sharp, and was perceived as making our characters look angry. This was successful in that many shots of the film required the protagonist to be very angry, but for our baseline look, we needed a softer touch. A third "dissipation" 2d style target helped bring a more gaseous and less aggressive look. Finally, lowering the relative frequency of the "plume" 2b image and replacing it with a more graphic shape, dubbed the "graphic ramp" 2c helped keep only the critical contributions at the desired scale. Taken together this became the NST look of our main character, and ultimately the larger population of fire characters. For moments when our fiery protagonist became more angry, we added controls for the animators to vary the contributions of our curated baseline and angry style image sets. We synchronized these controls with state changes in the underlying pyro simulation and shading to emphasize the emotional changes with visual cues.

4 OMNIVIEW

Our NST system became so successful in stylizing the look of our fire characters that it became necessary to consider also stylizing the environment fire and smoke simulations that characterized "Firetown" and "Fireland" sets in the film. Unlike our fire characters which we can afford to stylize per frame, environment effects often have much larger voxel counts. For efficiency, such effects need to be created once, and reused from multiple camera angles. TNST supports multiple stylization viewpoints (dubbed "omniview"), but handling all possible views can increase the computational cost by orders of magnitude, beyond what is practical. By training a convolutional neural network, omniview TNST could be approximated for similar data. Deploying required running voxels through this feed forward network, and was thus dubbed "Feed Forward Transport-Based Neural Style Transfer" or FFTNST [Aurand et al. 2022]. In practice we trained several models approximating our fire styles as well as some other graphic patterns to successfully stylize braziers, smoke stacks, and fiery motorcycle exhaust. Beyond view independence, FFTNST could also scale to much larger voxel counts since it can be applied in a tiled fashion.

5 CONCLUSIONS

Despite initial concerns about NST's scalability, we ended up using the technique on nearly every shot of *Elemental*. The optimized execution time of 6 GPU seconds a frame on average meant that most artists did not notice the cost as it ran after a far heavier simulation process. Tuning NST for our characters took considerable effort, but once dialed in, generalized surprisingly well across all sorts of simulation states and movements. Only a minimal number of shots required settings overrides. FFTNST lowered the expertise required since applying the pre-trained model eliminated the need to tune the many parameters, and we see future use of the technique relying more heavily on this approach. Research is ongoing to re-design FFTNST to train style independent models that allow new style targets, even from pre-trained networks.

REFERENCES

Joshua Aurand, Raphael Ortiz, Silvia Nauer, and Vinicius C. Azevedo. 2022. Efficient Neural Style Transfer for Volumetric Simulations. *ACM Trans. Graph.* 41, 6, Article 257 (nov 2022), 10 pages. https://doi.org/10.1145/3550454.3555517

Byungsoo Kim, Vinicius C. Azevedo, Markus Gross, and Barbara Solenthaler. 2019. Transport-Based Neural Style Transfer for Smoke Simulations. *ACM Trans. Graph.* 38, 6, Article 188 (nov 2019), 11 pages. https://doi.org/10.1145/3355089.3356560

Byungsoo Kim, Vinicius C. Azevedo, Markus Gross, and Barbara Solenthaler. 2020. Lagrangian Neural Style Transfer for Fluids. *ACM Trans. Graph.* 39, 4, Article 52 (aug 2020), 10 pages. https://doi.org/10.1145/3386569.3392473