

Transport-Based Neural Style Transfer for Smoke Simulations

BYUNGSOO KIM, ETH Zurich
VINICIUS C. AZEVEDO, ETH Zurich
MARKUS GROSS, ETH Zurich
BARBARA SOLENTHALER, ETH Zurich



Fig. 1. Volcanic smoke simulation. Left: stylized output by our transport-based neural style transfer; right: close-up views of low-resolution base input and ours with a cloud motif⁰ and a smoke exemplar ©Richard Roscoe via [Jamriška et al. 2015].

Artistically controlling fluids has always been a challenging task. Optimization techniques rely on approximating simulation states towards target velocity or density field configurations, which are often handcrafted by artists to indirectly control smoke dynamics. Patch synthesis techniques transfer image textures or simulation features to a target flow field. However, these are either limited to adding structural patterns or augmenting coarse flows with turbulent structures, and hence cannot capture the full spectrum of different styles and semantically complex structures. In this paper, we propose the first Transport-based Neural Style Transfer (TNST) algorithm for volumetric smoke data. Our method is able to transfer features from natural images to smoke simulations, enabling general content-aware manipulations ranging from simple patterns to intricate motifs. The proposed algorithm is physically inspired, since it computes the density transport from a source input smoke to a desired target configuration. Our transport-based approach

Authors' addresses: Byungsoo Kim, ETH Zurich, kimby@inf.ethz.ch; Vinicius C. Azevedo, ETH Zurich, vinicius.azevedo@inf.ethz.ch; Markus Gross, ETH Zurich, grossm@inf.ethz.ch; Barbara Solenthaler, ETH Zurich, solenthaler@inf.ethz.ch.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2019 Association for Computing Machinery.

0730-0301/2019/11-ART188 \$15

<https://doi.org/10.1145/3355089.3356560>

allows direct control over the divergence of the stylization velocity field by optimizing incompressible and irrotational potentials that transport smoke towards stylization. Temporal consistency is ensured by transporting and aligning subsequent stylized velocities, and 3D reconstructions are computed by seamlessly merging stylizations from different camera viewpoints.

CCS Concepts: • **Computing methodologies** → **Physical simulation**; **Neural networks**.

Additional Key Words and Phrases: physically-based animation, fluid simulation, neural style transfer

ACM Reference Format:

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 (November 2019), 11 pages. <https://doi.org/10.1145/3355089.3356560>

1 INTRODUCTION

Physically-based fluid simulations have become an essential part in digital content production. Due to the complexity of the underlying mathematical models, the process of manipulating fluids for simultaneously achieving controllable and realistic behavior, however, is tedious and time-consuming. Previous approaches [Inglis et al. 2017; Treuille et al. 2003] relied on optimization techniques to generate

⁰http://storage.googleapis.com/deepdream/visualz/tensorflow_inception/index.html

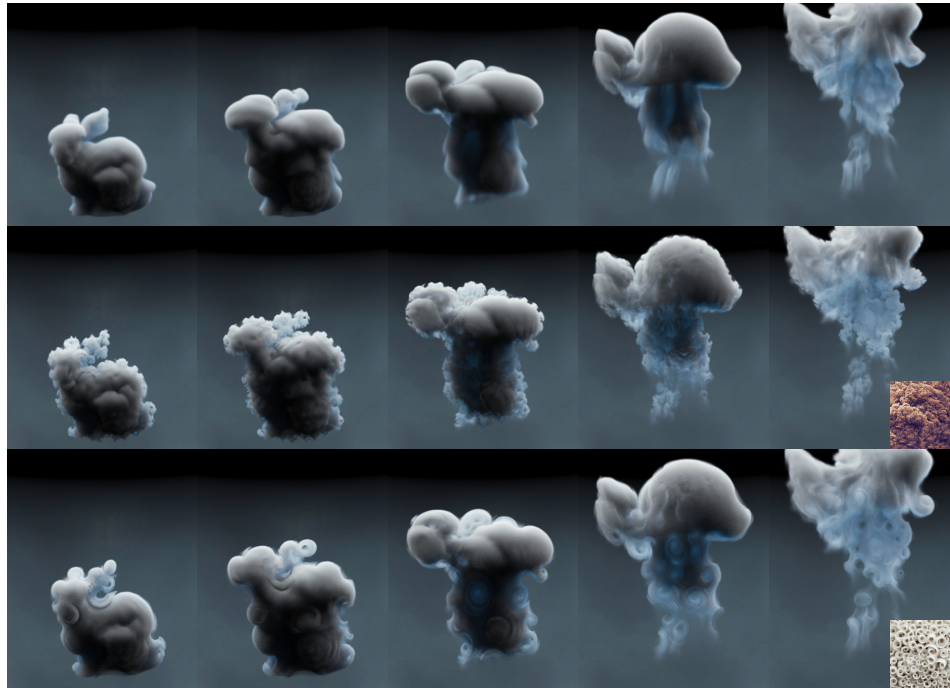


Fig. 2. Frames of a style transfer smoke example: base simulation (top), stylized output with volcano (middle, ©Richard Roscoe) and spiral¹ (bottom) images.

artificial forces in a flow solver to match user-designed keyframes for smoke [Treuille et al. 2003] and liquid animations [Nielsen and Bridson 2011]. Optimization techniques are computationally challenging since the space of control forces is naturally large [Pan and Manocha 2017], limiting the applicability of these methods to relatively coarse grid resolutions. Moreover, since these techniques rely on manually crafted keyframes, artistic manipulation is restricted to reproducing given 3D shapes in their entirety, while modification of small to medium scale flow features is not easily attainable.

Post-processing methods for fluids aim at enabling detailed feature control by patch-based texture and velocity synthesis. While current patch-based techniques focus on controlling structural patterns [Jamriška et al. 2015; Ma et al. 2009], they are limited to 2D flows. Velocity synthesis approaches allow augmentation of coarse simulations with turbulent structures [Kim et al. 2008; Sato et al. 2018], but cannot capture the full spectrum of different styles and complex semantics. Ideally, to support artistic manipulations of flow data, post-processing methods should enable multi-level control of flow features with automatic instantiation of patterns.

Inspired by Neural Style Transfer (NST) methods for images [Gatys et al. 2015] and meshes [Kato et al. 2018; Liu et al. 2018], we propose a novel method to synthesize semantic structures onto volumetric flow data by taking advantage of the simple yet powerful machinery developed for image editing. We modify 3D density fields by combining individual 2D stylizations from multiple views, which are synthesized by matching features of a pre-trained Convolutional Neural Network (CNN). Since the CNN is trained for image classification tasks, a vast library of patterns and class semantics is available, enabling novel content-aware flow manipulations that

range from transferring low (edges and patterns) to high (complex structures and shapes) level features from images to smoke simulations (Figures 1, 12 and 13). In this way, our method allows for automatic instantiation of structures in flow regions that naturally share features with a given target pattern or semantic class.

Crucially, and in contrast to existing NST methods, our style transfer algorithm is physically inspired. It computes a velocity field that stylizes a smoke density with an input target style, yielding results that naturally model the underlying transport phenomena. To improve temporal consistency, we propose a method which aligns stylization velocities from adjacent frames, enabling the control of how smoothly stylized structures change in time. To handle volumetric smoke stylization, multiple stylized 2D views are seamlessly combined into a 3D representation, resulting in coherent stylized smoke structures from arbitrary camera viewpoints. The proposed method is end-to-end differentiable and it can be readily optimized by gradient descent approaches. This is enabled by a novel volumetric differentiable smoke renderer, which is tailored for stylization purposes. Our results demonstrate that our method captures a wide spectrum of different styles and high-level semantics, and hence can be used to transfer patterns and regular structures, turbulence effects, shapes and artistic styles onto existing simulations.

2 RELATED WORK

Patch-based Appearance Transfer methods change the appearance of a source image or texture to a target by matching small compact regions called patches or neighborhoods. Kwatra et al. [2005]

¹http://ardezart.com/?attachment_id=252



Fig. 3. Stanford Bunny shaped smoke stylized with spiral patterns¹ for multiple views ($[-30^\circ, 30^\circ]$, every 15 degrees). Our method focuses the instantiation of patterns on smoke regions that share similarities with the target motif. Additionally, augmented flow structures change smoothly when the camera moves around the object.

employ local similarity measures to optimize an energy-based formulation, enabling the animation of texture patches by flow fields. Their approach was extended to liquid surfaces [Bargteil et al. 2006; Kwatra et al. 2006], and further improved by modifying the underlying texture based on visually salient features of the liquid mesh [Narain et al. 2007]. Bousseau et al. [2007] proposed a bidirectional advection scheme to reduce local patch distortions in a video watercolorization setup. Regenerative morphing and image melding techniques were combined with patch-based tracking to produce in-betweens for artist-stylized keyframes [Browning et al. 2014]. Jamriška et al. [2015] improved the temporal coherence aspect of previous energy-based formulations by reducing the wash-out effects that appear when textures are advected during long periods. Although patch-based appearance transfer methods were successful in synthesizing temporally coherent textures for flow animations, these are limited to 2D setups. For a broad discussion of patch-based texture synthesis works we refer to [Barnes and Zhang 2017].

Velocity Synthesis methods augment flow simulations with detailed flow fields to produce a desired effect. Due to the inability of numerical solvers to capture different energy scales of flow phenomena, sub-grid turbulence [Kim et al. 2008; Schechter and Bridson 2008] was modelled for increased realism. This was later extended to model turbulence in the wake of solid boundaries [Pfaff et al. 2009] and liquid surfaces [Kim et al. 2013]. Sato et al. [2018] transferred turbulence data from a source to a target scene similarly to patch-based appearance transfer methods: the target simulation is subdivided into smaller patches which are matched to the ones of the source simulation. Patterns are matched by the combination of patchwise weighted L2 distance functions on velocity and density data, performed in a two-level search. However, their method is limited to turbulent features only, and more general style transfer between distinct simulations is not demonstrated. Ma et al. [2009] synthesized velocity fields with example-based textures for artistic manipulations, but their method is limited to simple 2D patterns.

Fluid Control aims to define the overall shape and behavior through user-specified keyframes or reference images. Optimal [McNamara et al. 2004; Treuille et al. 2003] and proportional-derivative [Fattal and Lischinski 2004; Shi and Yu 2005] controllers define a set of forces that guide fluid simulation states to desired configurations. These methods were extended to match simulations from different resolutions [Nielsen et al. 2009], guide fluids [Nielsen and Bridson 2011; Rasmussen et al. 2004], volume-preserving morphing [Raveendran et al. 2012], improve performance [Pan and Manocha 2017],

and model more accurate boundary conditions while distinguishing low and high frequencies [Inglis et al. 2017]. However, due to the inherent high dimensionality of the configuration space of fluid solvers these methods are still computationally challenging, making detailed fluid control hard to achieve. Additionally, they require the specification of target shapes for control, and automatic stylization of fluid features is not possible. Guided simulation control can also be used to reconstruct target smoke images. Okabe et al. [2015] proposed an appearance transfer method for image-based 3D reconstruction of smoke volumes, while Eckert et al. [2018] utilized proximal operators to reconstruct both the fluid density and motion from single or multiple views.

Machine Learning & Fluids. Combining fluid simulation with machine learning was first demonstrated by Ladický et al. [2015]. The authors modeled a Lagrangian-based fluid solver by employing Regression forests to approximate particle positions and velocities given a neighborhood configuration. CNN-based architectures were employed in Eulerian contexts to substitute the pressure projection step [Tompson et al. 2017; Yang et al. 2016] and to synthesize flow simulations from a set of reduced parameters [Kim et al. 2019]. An LSTM-based [Wiewel et al. 2019] approach predicted changes on pressure fields for multiple subsequent time-steps. Closer to our work, Chu and Thuerey [2017] enhance simulations with patch correspondences between low and high resolution simulations. The patches are automatically created in a low resolution simulation, and then advected and deformed by the underlying flow field. A temporally coherent Generative Adversarial Network (GAN) was designed for smoke simulation super-resolution [Xie et al. 2018], removing the reliance on Lagrangian tracking of features of the previous approach. Their method can produce detailed, high-quality results, but it does not support transfer of different smoke styles.

Neural Style Transfer is the process of rendering image content in different styles by exploring CNNs. The seminal work of Gatys et al. [2015] was the first to transfer painting styles to natural images. Their model relies on extracting the content of an image by measuring filter responses of a pre-trained CNN, while modelling the style as summary feature statistics. The network's filter responses decompose the image complexity into multiple levels, ranging from low-level features to high-level semantics. Given a target style, NST approaches optimize CNN feature distributions of a source image style, while keeping its original content. Ruder et al. [2016] implemented style transfer for video sequences, addressing temporal

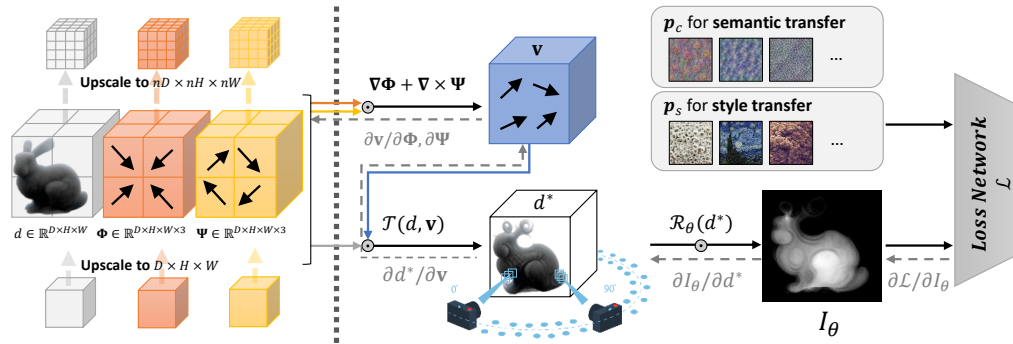


Fig. 4. Pipeline of our TNST method. On the left, three input fields d , Φ , Ψ for the stylization algorithm are shown. Φ , Ψ are iteratively updated during the optimization, while the stylized density output is represented by d^* . All fields are firstly downsampled and stylized on their coarser representations, so features can be enhanced through larger regions of the smoke. Cubic upsampling is performed on d , Φ , Ψ and the stylization runs again on a finer resolution; this process repeats until the specified resolution is matched. The right side of the diagram illustrates how our optimization works. The black arrows show the direction of a feed-forward pass from potentials Φ , Ψ to the loss network, and the gray arrows represent the backpropagation path for computing gradients.

coherency issues due to occluded regions and long term correspondences, while Mordvintsev et al. [2018] discuss the impact of different choices of image parameterizations for NST. For a detailed review of NST methods we refer to [Jing et al. 2019].

Differentiable rendering allows the computation of derivatives of image pixels with respect to the variables used for rendering the image, e.g., vertex positions, normals, colors, camera parameters, etc. These derivatives are crucial to optimization, inverse problems and deep learning backpropagation. Loper and Black [2014] proposed the first raster-based fully differentiable rendering engine with automatically computed derivatives. Anisotropic probing kernels were used to project 3D volumetric data similarly to x-ray scans [Qi et al. 2016]. Tulsiani et al. [2017] used a differentiable ray consistency approach to leverage different types of multi-view observations which can vary from depth and color to foreground masks and normals. Differentiable volume sampling was implemented by Yan et al. [2016] to obtain 2D silhouettes from 3D volumes, adopting a similar sampling strategy as spatial transformer networks [Jaderberg et al. 2015]. Kato et al. [2018] and Liu et al. [2018] proposed a raster-based differential rendering for meshes with approximate and analytic derivatives, respectively. Recently, there is a growing interest on differentiable ray marching. Li et al. [2018] introduces the first general-purpose differentiable ray tracer by removing discontinuities that appear when including visibility terms by directly sampling Dirac delta functions, while a differentiable path-tracer for inverse volumetric rendering with joint estimation of geometry was proposed by Velinov et al. [2018]. Unlike previous approaches, our proposed differentiable renderer is the first to specifically tackle volumetric data stylization.

3 TRANSPORT-BASED NEURAL STYLE TRANSFER

Our method employs pre-trained CNNs for natural image classification as both feature extractor and synthesizer. As an alternative, we considered CNNs trained on synthetic 3D representations such as voxels [Wu et al. 2015], meshes [Masci et al. 2015] and point clouds [Qi et al. 2017]. However, CNNs trained on 2D natural images have seen richer and denser information as there is a more

expressive incidence of high-frequency features [Qi et al. 2016], and data-sets have been thoroughly analyzed in terms of interpretability. Thus, as a classification CNN gets deeper, it shows its hierarchical interpretation of natural images organized from low-level patterns to high-level semantics [Olah et al. 2017].

The original neural style transfer (NST) [Gatys et al. 2015] transforms an initial noise image I to match the content (I_c) and style (I_s) of input target images. The content loss \mathcal{L}_c measures selected filter responses from a pre-trained classification CNN, while the style loss \mathcal{L}_s measures the difference between specific filter’s statistical distributions. The neural style transfer solves the optimization of

$$\hat{I} = \arg \min_I \alpha \mathcal{L}_c(I, I_c) + \beta \mathcal{L}_s(I, I_s), \quad (1)$$

where the weights α and β control how the content and style modify the initial image I along the optimization process.

Applying existing NST methods to stylize smoke data will lead to arbitrary creation of sources, since the volumetric density field is evaluated as an intensity image. Thus, we propose a transport-based neural style transfer (TNST), in which the stylization is driven by velocity fields instead of direct pixel / voxel corrections as introduced in Figure 4. The transport-based approach yields more degrees of freedom than directly changing the densities; specifically, it will yield a vector field while a standard value-based approach will output a scalar field correction. This is particularly useful in 3D, since the directional information encoded by the vector field will be used to merge stylizations from distinct camera viewpoints. Additionally, this approach enables the control over smoke density sources and sinks during stylization: we implement a divergence control through the decomposition of the stylization vector field into its incompressible and irrotational parts. A comparison between value- and velocity-based stylizations is shown in Figure 5.

3.1 Single-frame Multi-view Stylization

We define a single-frame loss for a given input image $I \in \mathbb{R}^{H \times W}$

$$\mathcal{L}(I, \mathbf{p}_c, \mathbf{p}_s) = \sigma [\alpha \mathcal{L}_c(I, \mathbf{p}_c) + \beta \mathcal{L}_s(I, \mathbf{p}_s)], \quad (2)$$

Table 1. Symbols, operators and configurable parameters

$\mathcal{L}_c, \mathcal{L}_s$	Content and style losses
α, β	Weights controlling content and style losses
d, \mathbf{u}	Input density and simulation velocity field
d^*, \mathbf{v}	Stylized density and stylization velocity field
σ	Density integrated spatially for a single frame
Φ, Ψ	Irrotational and incompressible potentials
λ	Weight between irrotational and incompressible vector fields
$\mathbf{p}_c, \mathbf{p}_s$	Content and style input parameters
\mathcal{R}, \mathcal{T}	Rendering and advection operators
$\mathcal{F}^l, \hat{\mathcal{F}}^l$	CNN's spatial and flattened feature maps for layer l
\mathcal{M}^l	User-defined feature map at layer l for semantic transfer
H, W, C	Height, width and channels of feature map or image
ω, w	Temporal coherency weight and window size
γ	Transmittance absorption factor
θ, Θ	Individual viewpoint and set of viewpoints
η	Learning rate size



Fig. 5. Value-based (left) against transport-based density optimization (right) with a flower motif⁰. The value-based approach used in traditional image stylization approaches produces ghosting artifacts and thinner smoke structures, since density sources can be created and removed to match targeted features.



Fig. 6. Results from semantic transfer of a net structure⁰. Irrotational (left), mixed (middle) and incompressible (right) velocity fields.

where $\sigma = \sum_i^H \sum_j^W I_{ij}$ is the per-pixel intensity ($I_{ij} \in [0, 1]$) integrated over the image, and \mathbf{p}_c and \mathbf{p}_s are user-specified parameters (Sections 3.2 and 3.3) that control content and style transfers. We normalize the loss function by the integrated pixel intensities, since, contrary to natural images, pixels from our rendered depiction represent smoke intensities that will be used for stylization.

Given the input density field $d : \mathbb{R}^3 \rightarrow \mathbb{R}$ and a velocity field $\mathbf{v} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, the transport function $\mathcal{T}(d, \mathbf{v})$ advects d by \mathbf{v} . Unlike image-based stylization, where pixels already contain color information of the represented image, our approach has to compute a valid

rendering of the flow data. The renderer $\mathcal{R}_\theta(d)$ outputs a grayscale image (Section 4) representing the density field for a specific view-point angle θ from a discrete set of viewpoints Θ . Our method optimizes a velocity field decomposed by a linear combination of its irrotational and incompressible parts by

$$\mathbf{v} = \lambda \nabla \Phi + (1 - \lambda) \nabla \times \Psi \quad (3)$$

to achieve a desired stylized density field by minimizing

$$\hat{\Phi}, \hat{\Psi} = \arg \min_{\Phi, \Psi} \sum_{\theta \in \Theta} \mathcal{L}(\mathcal{R}_\theta(d^*), \mathbf{p}_c, \mathbf{p}_s), \quad (4)$$

where $d^* = \mathcal{T}(d, \lambda \nabla \Phi + (1 - \lambda) \nabla \times \Psi)$ is the density field evolving towards stylization. Since our formulation optimizes for the scalar and vector potentials that transport the smoke, we allow the user to have direct control over the divergence of the stylization velocity field. Incompressible and irrotational velocity fields generate artistically different results and Figure 6 shows a comparison between both approaches. In order to create 3D structures, contributions from individual viewpoints \mathcal{R}_θ are summed, similarly to [Liu et al. 2018]. We will discuss camera view sampling and renderer specifications in Section 4. The next sections describe the loss functions that we use for semantic (\mathcal{L}_c) and style transfers (\mathcal{L}_s).

3.2 Semantic Transfer

Inspired by DeepDream², our method allows novel semantic transfer for stylizing smoke simulations by manipulating the content represented by the smoke. For example, smoke densities can be modified to portray patterns and shapes, such as squares or flowers, as depicted in Figure 12. Let $\mathcal{F}^l(I) \in \mathbb{R}^{H_l \times W_l \times C_l}$ denote a feature map of $[H_l, W_l]$ dimensions with C channels at the layer l of the network with respect to an input image I . The user-specified parameter \mathbf{p}_c consists of an array of feature maps $\mathcal{M} \in \mathbb{R}^{H_l \times W_l \times C_l}$ for all layers $l \in L$ specified by the array. We then define the content loss as to match features of a density field rendered image to a user-defined feature map by

$$\mathcal{L}_c(I, \mathbf{p}_c) = \sum_l \left[\frac{1}{H_l W_l C_l} \sum_i^{H_l} \sum_j^{W_l} \sum_k^{C_l} (\mathcal{F}_{ijk}^l(I) - \mathcal{M}_{ijk}^l)^2 \right], \quad (5)$$

where $\mathcal{F}_{ijk}^l(I)$ denotes an activated neuron of the CNN respective to the input image I at position (i, j) of the feature map's k^{th} channel. The feature map \mathcal{M} represents semantic features that will be transferred to the smoke (e.g., flowers), and it controls the abstraction level of structures created in the stylization process. Choosing a feature map that lies on deeper levels of the network will create more intricate motifs, as shown in Figure 12. The user can choose the abstraction level for the semantic transfer to match the specific content of an input image by selecting shallow levels of the network layers; or, conversely, match classification textual tags, enabling a stylization that maximizes tags (e.g., "volcano") on the output smoke.

Differently from previous image stylization approaches, we do not enforce the matching of content loss to the original unstyled image, and instead, the content loss is used to drive the flow data towards the creation of patterns. Since the smoke is modified by advecting its density towards stylization, we can guarantee that each

²<https://github.com/tensorflow/examples/tree/master/community/en/r1/deepdream.ipynb>



Fig. 7. Abstraction levels of style features and their impact on the smoke stylization result. We can control low (left), medium (center) and high (right) levels of features. The corner images show style representations corresponding to different feature levels.

iteration of the optimization will only slightly modify the original smoke by normalizing the gradients for updating the velocities with the fixed learning rate size (η).

3.3 Style Transfer

In addition to semantic transfer, which is designed to use "built-in" features of the pre-trained network *without* any reference image as input, our method allows the incorporation of a given input image style, as shown in Figure 13. The style is computed by correlations between different filter responses, where the expectation is taken over the spatial extension of the input image. Hence, in contrast to semantic transfer, we minimize the difference between feature distributions. Given $\hat{\mathcal{F}}_k^l(I)$, which is the flattened one-dimensional version of a 2D filter map at the k^{th} channel, the Gram matrix entry for two channels m and n is

$$G_{mn}^l(I) = \sum_i^{H_l \times W_l} \hat{\mathcal{F}}_{mi}^l(I) \hat{\mathcal{F}}_{ni}^l(I), \quad (6)$$

where i iterates over all pixels of the vectorized filter. Thus, the Gram $G^l(I)$ matrix of a l^{th} layer has dimensions $C_l \times C_l$. The Gram matrix computes the dot product between all filter responses from a layer, storing correspondences of channels denoted by the row and column of an entry. The user-specified parameter \mathbf{p}_s consists of a target image I_s and a set of layers for which the style will be optimized for. Thus, the normalized loss function \mathcal{L}_s for matching styles between an input image and a target style image is

$$\mathcal{L}_s(I, \mathbf{p}_s) = \sum_l \left[\frac{1}{4C_l^2(H_l \times W_l)^2} \sum_{m,n}^{C_l} \left(G_{mn}^l(I) - G_{mn}^l(I_s) \right)^2 \right]. \quad (7)$$

Similarly to our semantic transfer, the Gram matrix layer choice in Equation (7) controls different abstraction levels of the stylization, as illustrated in Figure 7. However, the style transfer does not match features that have spatial correlations relative to the input image, but rather approximates filter response statistics. We further highlight differences between semantic and style transfer in Section 5.1.

3.4 Time-Coherent Stylization

As densities are updated with the simulation advancement, distinct features can be emphasized by semantic and style transfer losses over different frames. Thus, flickering will occur if time-coherency between frames is not enforced explicitly, as shown in Figure 8. Given that the velocities of the original simulation transport densities over time, we use them to align stylization velocities

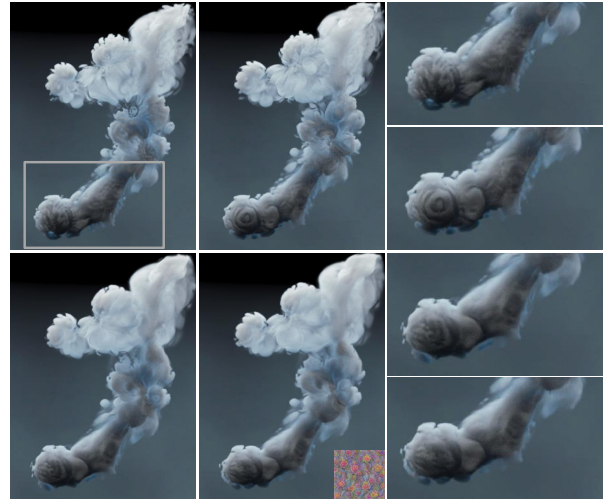


Fig. 8. Two subsequent frames of the smoke jet example stylized with a flower⁰ motif, no time coherence (top) and window size 9 (bottom). For each frame, a close-up view corresponding to the highlighted region is shown on the right. Using our algorithm with a bigger window size ensures that the structures created in one frame are propagated to subsequent stylizations.

computed independently for different frames. Once these velocities are aligned, we update a single frame stylization velocity field by smoothing subsequent aligned velocities together. Specifically, we define $\mathbf{U} = \{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{n-1}, \mathbf{u}_n\}$ as the set of simulation velocities computed for the whole simulation duration. The advection function \mathcal{T}_i^j that takes a stylization velocity at the i^{th} frame to the j^{th} frame is

$$\mathcal{T}_i^j(\mathbf{v}_i, \mathbf{U}) = \begin{cases} \mathcal{T}(\dots \mathcal{T}(\mathcal{T}(\mathbf{v}_i, \mathbf{u}_i), \mathbf{u}_{i+1}) \dots, \mathbf{u}_{j-1}), & \text{if } i < j \\ \mathcal{T}(\dots \mathcal{T}(\mathcal{T}(\mathbf{v}_i, -\mathbf{u}_{i-1}), -\mathbf{u}_{i-2}) \dots, -\mathbf{u}_j), & \text{if } i > j \\ \mathbf{v}_i, & \text{if } i = j \end{cases} \quad (8)$$

where \mathcal{T} is a function that advects a velocity or a density field for a single time-step. Equation (8) is recursive, and aligning a velocity field defined n frames away from a specific frame requires n evaluations of the advection function. A temporally coherent velocity for stylization of frame t is given as a linear combination of aligned neighbor velocity fields

$$\mathbf{v}_t^* = \sum_{i=t-w}^{t+w} \omega_i \mathcal{T}_i^t(\mathbf{v}_i, \mathbf{U}), \quad (9)$$

where w is the number of neighboring frames evaluated in time, $2w + 1$ is the window size and ω_i is a weighting term. Let $\mathbf{V}_t = \{\mathbf{v}_{t-w}, \mathbf{v}_{t-(w-1)}, \dots, \mathbf{v}_t, \dots, \mathbf{v}_{t+(w-1)}, \mathbf{v}_{t+w}\}$ be the window of stylization velocities at time t obtained by the combination of corresponding potential windows Φ_t, Ψ_t defined in a range from $t - w$ to $t + w$. The time-coherent multi-view stylization optimization is

$$\hat{\Phi}_t, \hat{\Psi}_t = \arg \min_{\Phi_t, \Psi_t} \sum_{i=t-w}^{t+w} \sum_{\theta \in \Theta} \mathcal{L}(\mathcal{R}_\theta(\mathcal{T}(d_i, \mathbf{v}_i^*)), \mathbf{p}_c, \mathbf{p}_s). \quad (10)$$

In practice, evaluating directly Equation (10) becomes infeasible as the number of neighbors increases. The memory used by the

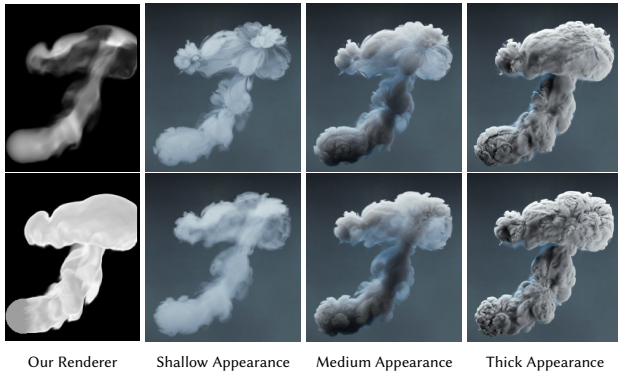


Fig. 9. The value of γ controls how the smoke density is stylized. Smoke images (left) produced by our renderer with $\gamma = 0.01$ (top) and $\gamma = 1$ (bottom). The final smoke renderer is configured with varying thickness, highlighting how the stylization gets transferred for different smoke appearances.

automatic differentiation procedure to compute derivatives quickly grows as the window size increases. Thus, we approximate the solution of Equation (10) by first evaluating Equation (4) to find a set of stylization velocities computed for a single frame. Then, we merge the velocities per-frame individually using Equation (9). This is performed iteratively for all simulation frames of a sequence, and the multi-view time-coherent process is summarized in Algorithm (1).

Algorithm 1 Multi-view Time-coherent Smoke Stylization

```

while  $i < n_{iter}$  do
  while  $t < n_{frames}$  do
    Compute density by  $d_t^* = \mathcal{T}(d_t, \mathbf{v}_t^*)$ 
    for  $\theta \in \Theta$  do
      Render  $I_t^\theta = \mathcal{R}_\theta(d_t^*)$  with angle  $\theta$ 
      Obtain  $\nabla\Phi_t^\theta, \nabla\Psi_t^\theta$  from  $\mathcal{L}(I_t^\theta, \mathbf{p}_c, \mathbf{p}_s)$ 
    end for
    Merge gradients from views  $\nabla\Phi_t^\theta, \nabla\Psi_t^\theta$  to obtain  $\nabla\Phi_t^*, \nabla\Psi_t^*$ 
    for  $t_w = t - w, t_w < t + w$  do
      Align gradients  $\nabla\Phi_t^*, \nabla\Psi_t^*$  to obtain  $\nabla\Phi_t, \nabla\Psi_t$  Eq. 8
    end for
     $\Phi_t = \Phi_t + \eta \nabla\Phi_t, \Psi_t = \Psi_t + \eta \nabla\Psi_t$ 
     $\mathbf{v}_t^* = \lambda \nabla\Phi_t + (1 - \lambda) \nabla \times \Psi_t$ 
  end while
end while
    
```

4 DIFFERENTIABLE SMOKE RENDERER

Similar to the flat shading approach proposed by Liu et al. [2018] for stylizing meshes, our smoke renderer is lightweight. The optimization of Equation (4) heavily relies on rendered density representations, and an overly sophisticated volumetric renderer compromises efficiency. Our renderer outputs grayscale images, in which pixel intensity values will correspond to density occupancy data. Thus, modelling smoke self-shadowing would map shadowed regions to empty voxels on the rendered image. Nevertheless, our results show that meaningful correspondences between the stylization velocities and density fields can be computed on representations that do not match perfectly the ones produced by the final rendered image.

The smoke stylization optimization usually performs many iterations, computing derivatives of the loss function (Equation (4)) with respect to the velocity field by automatic differentiation. Therefore, the volumetric rendering requires efficiency. Our lightweight differentiable rendering algorithm only incorporates a single directional light traced directly from the pixel rendered from an orthographic camera. We measure how much of this single light ray gets transmitted through the inhomogeneous participating media, which is described by [Fong et al. 2017], to compute the transmittance and the image pixel grayscale value as

$$\tau(\mathbf{x}, \mathbf{r}) = e^{-\gamma \int_{\mathbf{x}}^{\mathbf{r}^{max}} d(\mathbf{r}) d\mathbf{r}}$$

$$I_{ij} = \int_0^{\mathbf{r}^{max}} d(\mathbf{x}) \tau(\mathbf{x}, \mathbf{r}) d\mathbf{x}, \quad (11)$$

where \mathbf{r}_{ij} is a vector traced from pixel ij into the normal direction of an orthographic camera, $d(\mathbf{r})$ is the density value, γ is a transmittance absorption factor, and r_{max} is the maximum length of the traced ray. The value computed at each image pixel is the integral of the transmittance multiplied by the density values, mapping empty and full smoke voxels to 0 and 1, respectively. We additionally multiply the transmittance and densities along the integration ray since it generates richer features for thicker smoke scenarios. Comparisons between this approach against simply integrating the transmittance along the view-ray are shown in our supplementary material.

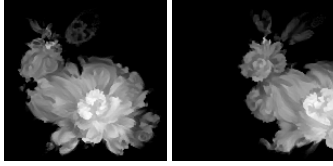
The smoke density is linearly mapped to extinction using the scaling factor γ , which determines how quickly light gets absorbed by the smoke. Figure 9 shows that minimizing the discrepancy between the final rendered smoke and the representation in which it is optimized is important. Setting low transmittance constants in the stylization renderer will result in more aggressive smoke modifications towards the normal view direction, while high transmittance will over-constrain the stylization velocity field to the smoke surface that is closer to the camera. Figure 9 shows the effect of varying γ values with the final rendered smoke thickness: low γ values will produce better results for shallow smoke appearances, while higher γ values will more efficiently stylize thicker smoke.

4.1 Camera Design Specifications

Participating media naturally incorporates transparency, and a single-view stylization update will be propagated inside the volumetric smoke even though the rendered image is two dimensional. Therefore it is not necessary to uniformly cover every viewpoint of the smoke with equal probability as in [Liu et al. 2018]. Given a predefined camera path, we use Poisson sampling around a small area of its trajectory (Figure 11, left) to avoid bias that would be introduced by a fixed set of viewpoints.

Since feature maps obtained from 2D views of the camera are used, we specify that the image rendered by the camera is invariant to zooming, panning and rolling. This means that if the camera is moving (as in Figure 11, left), our renderer automatically centers the smoke representation in the frame, only responding to variations of the viewing angle. These invariances ensure that filter map activations remain constant as long as no new voxels are shown in the rendered image; rotations, however, have to be accounted for. Thus,

our renderer camera position is parameterized by the polar coordinates tuple $\theta = (\theta_1, \theta_2)$, while the camera always points to a fixed point inside the smoke. Note that this simplification is only possible since we are adopting an orthogonal camera, and a perspective projection might reveal new voxels with translational movement. The inset image shows how enhanced features (e.g., patterns at the bunny face) vary due to translations in the image space, in which the stylization should remain constant.



In order to evaluate Equation (11) for multiple perspectives, we need to integrate smoke voxels along the camera view direction. Implementing a classic ray-marching sampling along an arbitrary ray direction is challenging in Deep Learning frameworks, which are usually optimized for tensor operations. Thus, we adopted the spatial transformer network (STN) of Jaderberg et al. [2015]. The STN instances a rotated 3D domain with the same dimensionality as the original one that is aligned with the camera view as illustrated in Figure 11 (right). This allows us to evaluate samples by evoking simple built-in features that implement voxel summations along the view direction of each pixels' ray. These specifications make our rendering algorithm efficient, accounting for about 30% of time taken for processing a batch (see Table (2)).

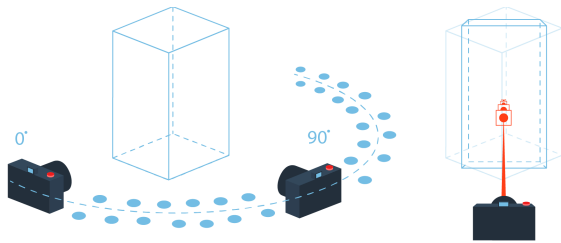


Fig. 11. Multiview camera configuration. We sample a camera path with Poisson sampling, which prevents smoothing of density details between predefined viewpoints (left). The volumetric smoke grid is aligned with the camera viewpoint to facilitate light ray integration (right).

5 RESULTS

We demonstrate that our approach can reliably transfer various styles from images onto volumetric flow data, with automatic semantic instantiation of features and artistic style transfer. All our stylization examples employed a mask with soft edges that is extracted from the original smoke data. The mask is applied to the potential field, and it restricts modifications to be close to the original smoke border while enhancing temporal accuracy near smoke boundaries (Section 5.2). We provide further results, extended masking discussion, influence of parameters and additional 2-D examples in our supplemental material. The advection operator \mathcal{T} is implemented by the McCormack method [Selle et al. 2008]. We refer the reader to the supplemental video for the corresponding animations.

Equations 5 and 7 are optimized by stochastic gradient descent, with the gradients computed by backpropagation on GoogleNet [Szegedy et al. 2015]. Although we use automatic differentiation, analytic differentiation would allow us to fit even bigger simulation

examples [Liu et al. 2018]. We modified the original stride size of GoogleNet's first layer from two to one³ to remove checkerboard patterns that occur when the kernel size is not divisible by the CNN's stride size. We use a fixed learning rate and apply multiscale stylization and Laplacian pyramid gradient normalization techniques² for boosting lower frequencies. Parameters and performance values for all examples are summarized in Table (2). Our implementation uses *tensorflow* evaluated on a TITAN Xp GPU (12GB). The input simulations have been computed with different solvers. We used *mantaflow* for the smokejet and bunny examples in Figure 12 and Figure 13, Houdini for computing the volcano in Figure 1, and a dataset from Sato et al. [2018] in Figure 14.

5.1 Semantic and Style Transfers

To demonstrate how our method performs under distinct style and semantic transfers, we designed two instances of buoyancy-driven smoke: a smokejet with a sphere-shaped source and an initial horizontal velocity, and a smoke initialized with the Stanford bunny shape (Figure 12, left). For all examples shown in Figures 12 and 13 we used 20 iterations for each scale with a learning rate of 0.002, 3 Laplacian subdivisions and 9 camera views for a single frame Poisson sampled around the original view with elevation (θ_1) and azimuth (θ_2) ranges spanning $[-5^\circ, 5^\circ]$ and $[-10^\circ, 10^\circ]$ respectively. The stylized examples show that our method is able to augment the original flow structures of the smoke, generating a wide set of artistic and natural 3D effects.

Examples in Figure 12 demonstrate results obtained by applying the semantic style transfer loss from Equation (5). All the feature maps are from the GoogleNet [Szegedy et al. 2015] architecture; the captions indicate which layer of the CNN is used for the optimization. Activating different layers of the network will yield stylization results which will vary in the complexity of instantiated structures. In the two first examples we used filters closer to initial layers, which depict simpler patterns that occur at lower levels of abstraction. These patterns are used by higher levels to composite more complex structures. As the layers become deeper, the network is able to produce more intricate motifs, such as structures similar to flowers, fur, or ribbons.

Examples shown in Figure 13 apply the style loss of Equation (7). To demonstrate the flexibility of our approach, we used three different image categories for testing the style transfer loss: photorealistic (first and second columns), artistic (third and fourth columns) and patterns (fifth and sixth columns). For all these examples, a mix of convolution layers from different levels of the CNN is used, similarly to [Gatys et al. 2015]. The representations of the layers employed on the style transfer are depicted below each input style image. Figure 2 shows distinct frames of the bunny-shaped rising smoke stylized with the volcano and spiral input images from Figure 13. In Figure 14 we compare our results with the example-based turbulence transfer method of Sato et al. [2018]. The style of the single frame rendering (Figure 14, middle) of their method output is transferred by our approach to an input coarse simulation. The results show that our

³<https://medium.com/mlreview/getting-inception-architectures-to-work-with-style-transfer-767d53475bf8>

approach is able to generate similarly detailed flow structures with only a given reference image.

Besides individually transferring semantics and styles, our method is also flexible to allow the combination of these techniques. An example in the accompanying video demonstrates the semantic instantiation of cloud motifs merged with the style of a fire texture, while a similar stylization shows ribbon patterns combined with Starry Night painting style. Figure 1 shows a low-resolution volcanic setup in which a combination of cloud motifs and a volcano texture was used to create turbulent details. The thick smoke produced by volcanic ashes poses challenges to our renderer, since it quickly saturates transmittance values. Nevertheless, our renderer is able to create structures that consistently correlate with the input smoke.

Our method is also able to control the amount of smoke dissipation by the decomposition of the stylization velocity field into its incompressible and irrotational parts. Figure 6 and examples in the supplemental video show that different artistic patterns are created depending on the constraints imposed on the velocity field.

Although our method is based on 2D representations of the smoke data, it can reliably cover multiple viewing directions without introducing bias towards certain views. This is allowed by the Poisson sampling of positions along the camera trajectory. Figure 3 shows the bunny smoke example stylized with a spiral pattern from different viewpoints. Transferred structures change smoothly when the camera moves around the object. An example in the accompanying video compares results of the smoke bunny example stylized with single view and multiple camera views. The multiple camera views are Poisson sampled with 180 degrees angle range around the y -axis. The multi view approach shows more salient 3D structures when compared with the single view stylization.

Lastly, Figure 8 shows the impact of different time coherency window sizes (Equation (9)). We compare a window size of 1 (frame-based stylization) with a window size of 9 (used in all other examples) for multiple subsequent frames. Features heavily flicker with a small window size, while augmented structures change smoothly with larger windows. These results are better visualized in the accompanying video, in which we also included a comparison with a window size of 5.

5.2 Discussion

Differentiability. Note that all operators including advection and rendering require differentiability with respect to the velocities, so efficient gradient-based optimization methods can be employed. Traditional NST works by differentiating the loss of a classification network with respect to the image input, computing gradients of filter responses to image variations. Since classification networks convolve images to create filter responses, these filters are assumed to smoothly change with respect to image variations, and thus NST works without differentiability issues. This is the same for our work, however we additionally require that both the transport towards stylization and the smoke rendering to be differentiable. The rendering scheme denoted by Equation (11) is clearly differentiable. The MacCormack advection uses the Semi-Lagrangian method as its building blocks to correct error estimations. The correction is

differentiable and the Semi-Lagrangian algorithm works by sampling densities in previous positions. Thus, two components need to be considered for differentiation to work: estimation of particle trajectories and density sampling. The estimation of the particle trajectories is a linear ODE, and thus differentiable. Densities are estimated by grid sampling, and as shown in [Jaderberg et al. 2015] this is also differentiable when using linear interpolation kernels.

Performance and memory limitations. Table (2) shows the average time for stylizing a single frame of different simulation resolutions, with grids up to $200 \times 300 \times 200$. Performance was not the focus of this work, and as shown by extensive follow-up works to image-based NST [Ulyanov et al. 2016], we believe that real-time stylizations can be obtained by training networks to directly output stylized results. For higher resolutions, the limiting factor is the single GPU memory used for computing the backpropagation with Tensorflow automatic differentiation. As in [Liu et al. 2018], the memory limitation could be greatly reduced by using analytic differentiation.

Temporal Coherency and Boundary conditions. Features are instantiated by evaluating smoke representations independently for each frame. Our temporal coherency algorithm aligns and blends the creation of those features; however, due to the nature of the underlying physical phenomena, smoke structures might appear and disappear as the simulation advances. This might induce abrupt changes in the stylized smoke, specially when considering smoke edges. We did not post-process our results in order to have a fair evaluation, but this effect can be controlled by blending the results with the original smoke simulation or by a more aggressive masking scheme. Regarding boundary conditions, the final stylized smoke can only slightly penetrate objects inside the simulation, since it starts from a density field configuration that is already boundary respecting. We include stylization experiments for scenes that include obstacles in our supplementary material. *arterial.*

6 CONCLUSIONS

In this work, we presented the first transport-based Neural Style Transfer algorithm for smoke simulations. Our method enables automatic instantiation of a vast set of motifs through semantic transfer, which enables novel artistic manipulations for fluid simulation data. Additionally, the proposed method successfully synthesizes various different styles of input images, ranging from artistic to photorealistic examples. Even though 2D CNNs are employed, our differentiable renderer allows the creation of 3D volumetric structures from small set of views. Our stylization algorithm is able to handle high resolutions simulations up to 16 million voxels.

We are not aware of any other methods that use a volumetric differentiable rendering for optimizing 3D smoke data, and we believe that our work may inspire further research in this direction. For example, our differentiable renderer could be employed for reconstructing 3D smoke volumes from images as in [Eckert et al. 2018], and be extended to transfer image-based filters as in [Liu et al. 2018]. As further extensions, super-resolution can be thought as a specific type of style transfer [Johnson et al. 2016], and we believe that our work can be tailored towards improving current super-resolution methods for fluids [Xie et al. 2018]. Additionally, the style transfer

⁴<https://github.com/titu1994/Neural-Style-Transfer>

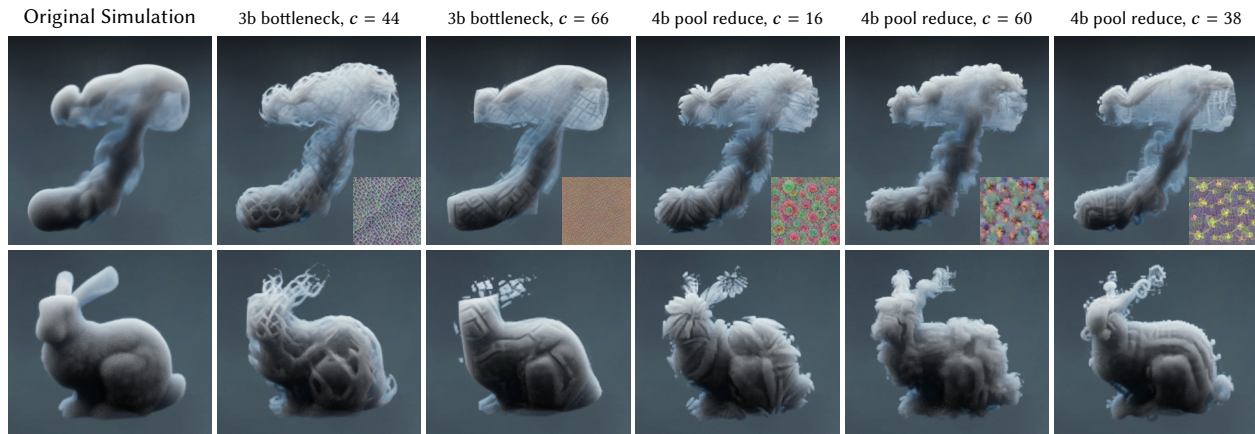


Fig. 12. Semantic transfer applied to a smokejet and bunny simulations (leftmost column). Images on the same column are stylized with the feature map depicted on the right corner⁰. The examples for semantic transfer depict different levels of abstraction, showing patterns that occur at shallow levels of the network (first two columns) and intricate motifs that are represented at deeper levels (last three columns).



Fig. 13. Style transfer applied to a smokejet and bunny simulations. We used photorealistic (first two columns), artistic (middle two columns) and pattern-based (last two columns) input images⁴ as input to the stylization algorithm. Fire exemplar ©Bunzellisa via pixabay.

Table 2. Parameters and performance statistics. We used a constant multi-scaling factor of 1.8, and the input size is firstly down-sampled to $61 \times 92 \times 61$ and up-scaled to $111 \times 166 \times 111$ and $200 \times 300 \times 200$. Computation time per frame includes all input scales.

Scene	Simulation		Learning Rate	Extinction Factor α	Multi Scale	# Target Layers	Computation Time per Frame (m)
	Resolution	# Frames					
Semantic Transfer (Fig. 12)	$200 \times 300 \times 200$	120	0.002	0.1	3	1	13.47
Style Transfer (Fig. 13)	$200 \times 300 \times 200$	120	0.002	0.1	2	3	12.68
Volcano (Fig. 1)	$200 \times 300 \times 200$	140	0.003	10	2	2	12.97
Sato et al. [2018] (Fig. 14)	$192 \times 256 \times 192$	140	0.005	5	2	3	11.97

quality could be improved by histogram normalization [Risser et al. 2017]. Our method does currently not handle color information to enable appearance transfer effects as shown in [Jamriška et al. 2015], which could be an interesting direction for further research.

ACKNOWLEDGMENTS

The authors would like to thank Fraser Rothnie for his artistic contributions. The work was supported by the Swiss National Science Foundation under Grant No.: 200021_168997.

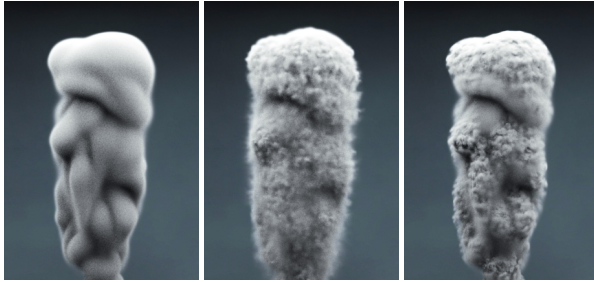


Fig. 14. Style Transfer comparison. From left to right: input density field of Sato et al. [2018], the result of applying the method of Sato et al., and our style transfer result using the middle image as stylization input.

REFERENCES

- Adam W Bargeil, Funshing Sin, Jonathan E Michaels, Tolga G Goktekin, and James F O'Brien. 2006. A Texture Synthesis Method for Liquid Animations. In *Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. 345–351.
- Connelly Barnes and Fang-Lue Zhang. 2017. A survey of the state-of-the-art in patch-based synthesis. *Computational Visual Media* 3, 1 (2017), 3–20.
- Adrien Bousseau, Fabrice Neyret, Joëlle Thollot, and David Salesin. 2007. Video warcolorization using bidirectional texture advection. *ACM ToG* 26, 3 (2007), 104.
- Mark Browning, Connelly Barnes, Samantha Ritter, and Adam Finkelstein. 2014. Stylized keyframe animation of fluid simulations. In *Proceedings of NPAR*. 63–70.
- Mengyu Chu and Nils Thuerey. 2017. Data-driven synthesis of smoke flows with CNN-based feature descriptors. *ACM ToG* 36, 4 (2017), 1–14.
- Marie-Lena Eckert, Wolfgang Heidrich, and Nils Thuerey. 2018. Coupled Fluid Density and Motion from Single Views. *Computer Graphics Forum* (2018).
- Raanan Fattal and Dani Lischinski. 2004. Target-driven smoke animation. In *ACM SIGGRAPH 2004*. 441.
- Julian Fong, Magnus Wrenninge, Christopher Kulla, and Ralf Habel. 2017. Production volume rendering. In *ACM SIGGRAPH 2017 Courses*. 1–79.
- Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2015. A neural algorithm of artistic style. *Nature Communications* (2015).
- Tiffany Inglis, Marie-Lena Eckert, James Gregson, and Nils Thuerey. 2017. Primal-Dual Optimization for Fluids. *Computer Graphics Forum* 36, 8 (2017), 354–368.
- Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. 2015. Spatial Transformer Networks. In *Proc. of the NIPS*. 2017–2025.
- Ondřej Jamriška, Jakub Fišer, Paul Asente, Jingwan Lu, Eli Shechtman, and Daniel Šykora. 2015. LazyFluids: appearance transfer for fluid animations. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 92.
- Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. 2019. Neural style transfer: A review. *IEEE TVCG* (2019).
- Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*.
- Hiroharu Kato, Yoshitaka Ushiku, and Tatsuya Harada. 2018. Neural 3d mesh renderer. In *Proceedings of the IEEE Conference on CVPR*. 3907–3916.
- Byungsoo Kim, Vinicius C. Azevedo, Nils Thuerey, Theodore Kim, Markus Gross, and Barbara Solenthaler. 2019. Deep Fluids: A Generative Network for Parameterized Fluid Simulations. *Computer Graphics Forum (Proc. Eurographics)* 38, 2 (2019).
- Theodore Kim, Jerry Tessendorf, and Nils Thuerey. 2013. Closest point turbulence for liquid surfaces. *ACM ToG* 32, 2 (2013), 15.
- Theodore Kim, Nils Thürey, Doug James, and Markus Gross. 2008. Wavelet turbulence for fluid simulation. In *ACM ToG*, Vol. 27. ACM, 50.
- Vivek Kwatra, David Adalsteinsson, Nipun Kwatra, Mark Carlson, and Ming C. Lin. 2006. Texturing fluids. In *ACM SIGGRAPH 2006 Sketches*. 63.
- Vivek Kwatra, Irfan Essa, Aaron Bobick, and Nipun Kwatra. 2005. Texture optimization for example-based synthesis. In *ACM SIGGRAPH 2005 Papers*. 795.
- L'ubor Ladický, SoHyeon Jeong, Barbara Solenthaler, Marc Pollefeys, and Markus Gross. 2015. Data-driven fluid simulations using regression forests. *ACM ToG* 34, 6 (2015).
- Tzu-Mao Li, Miika Aittala, Frédo Durand, and Jaakko Lehtinen. 2018. Differentiable Monte Carlo ray tracing through edge sampling. In *SIGGRAPH Asia 2018*. 1–11.
- Hsueh-Ti Derek Liu, Michael Tao, and Alec Jacobson. 2018. Paparazzi: Surface Editing by way of Multi-View Image Processing. *ACM ToG* (2018).
- Matthew M. Loper and Michael J. Black. 2014. OpenDR: An Approximate Differentiable Renderer. In *Computer Vision – ECCV 2014*, Vol. 8695. 154–169.
- Chongyang Ma, Li-Yi Wei, Baining Guo, and Kun Zhou. 2009. Motion field texture synthesis. In *ACM ToG*, Vol. 28. 110.
- Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. 2015. Geodesic convolutional neural networks on riemannian manifolds. In *Proceedings of the IEEE international conference on computer vision workshops*. 37–45.
- Antoine McNamara, Adrien Treuille, Zoran Popović, and Jos Stam. 2004. Fluid control using the adjoint method. In *ACM SIGGRAPH 2004*. 449.
- Alexander Mordvintsev, Nicola Pezzotti, Ludwig Schubert, and Chris Olah. 2018. Differentiable Image Parameterizations. *Distill* (2018). <https://doi.org/10.23915/distill.00012>
- Rahul Narain, Vivek Kwatra, Huai-Ping Lee, Theodore Kim, Mark Carlson, and Ming C Lin. 2007. Feature-guided Dynamic Texture Synthesis on Continuous Flows. In *Proceedings of the 18th Eurographics Conference on Rendering Techniques*. 361–370.
- Michael B. Nielsen and Robert Bridson. 2011. Guide shapes for high resolution naturalistic liquid simulation. In *ACM SIGGRAPH 2011*. 1.
- Michael B. Nielsen, Brian B. Christensen, Nafees Bin Zafar, Doug Roble, and Ken Museth. 2009. Guiding of smoke animations through variational coupling of simulations at different resolutions. In *Proceedings ACM SIGGRAPH/Eurographics SCA*. 217.
- Makoto Okabe, Yoshinori Dobashi, Ken Anjyo, and Rikio Onai. 2015. Fluid volume modeling from sparse multi-view images by appearance transfer. *ACM ToG* 34, 4 (2015), 93.
- Chris Olah, Alexander Mordvintsev, and Ludwig Schubert. 2017. Feature Visualization. *Distill* (2017). <https://doi.org/10.23915/distill.00007>
- Zherong Pan and Dinesh Manocha. 2017. Efficient Solver for Spacetime Control of Smoke. *ACM Trans. Graph.* 36, 4, Article 68a (July 2017).
- Tobias Pfaff, Nils Thuerey, Andrew Selle, and Markus Gross. 2009. Synthetic turbulence using artificial boundary layers. *ACM ToG* 28, 5 (2009), 1.
- Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. 2017. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 652–660.
- Charles R Qi, Hao Su, Matthias Nießner, Angela Dai, Mengyuan Yan, and Leonidas J Guibas. 2016. Volumetric and multi-view cnns for object classification on 3d data. In *Proceedings of the IEEE conference on CVPR*. 5648–5656.
- N. Rasmussen, D. Enright, D. Nguyen, S. Marino, N. Sumner, W. Geiger, S. Hoon, and R. Fedkiw. 2004. Directable photorealistic liquids. In *Proceedings of the ACM SIGGRAPH/Eurographics symposium on Computer animation*. 193.
- Karthik Raveendran, Nils Thuerey, Chris Wojtan, and Greg Turk. 2012. Controlling Liquids Using Meshes. In *Proc. of the ACM SIGGRAPH/Eurographics SCA*. 255–264.
- Eric Risser, Pierre Wilnot, and Connelly Barnes. 2017. Stable and Controllable Neural Texture Synthesis and Style Transfer Using Histogram Losses. (jan 2017). arXiv:1701.08893 <http://arxiv.org/abs/1701.08893>
- Manuel Ruder, Alexey Dosovitskiy, and Thomas Brox. 2016. Artistic style transfer for videos. In *German Conference on Pattern Recognition*. Springer, 26–36.
- Syuhui Sato, Yoshinori Dobashi, Theodore Kim, and Tomoyuki Nishita. 2018. Example-based turbulence style transfer. *ACM ToG* 37, 4 (2018), 84.
- Hagit Schechter and Robert Bridson. 2008. Evolving Sub-Grid Turbulence for Smoke Animation. In *In Proceedings of ACM Siggraph / Eurographics SCA*.
- Andrew Selle, Ronald Fedkiw, ByungMoon Kim, Yingjie Liu, and Jarek Rossignac. 2008. An Unconditionally Stable MacCormack Method. *Journal of Scientific Computing* 35, 2-3 (2008), 350–371.
- Lin Shi and Yizhou Yu. 2005. Taming liquids for rapidly changing targets. In *Proceedings of ACM SIGGRAPH/Eurographics symposium on Computer animation*. 229.
- Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going Deeper with Convolutions. In *Computer Vision and Pattern Recognition (CVPR)*.
- Jonathan Tompson, Kristofer Schlachter, Pablo Sprechmann, and Ken Perlin. 2017. Accelerating eulerian fluid simulation with convolutional networks. In *Proceedings of the 34th ICML Vol. 70*. JMLR. org, 3424–3433.
- Adrien Treuille, Antoine McNamara, Zoran Popović, and Jos Stam. 2003. Keyframe control of smoke simulations. *ACM ToG* 22, 3 (2003), 716.
- Shubham Tulsiani, Tinghui Zhou, Alexei A Efros, and Jitendra Malik. 2017. Multi-view supervision for single-view reconstruction via differentiable ray consistency. In *Proceedings of the IEEE conference on CVPR*. 2626–2634.
- Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, and Victor Lempitsky. 2016. Texture Networks: Feed-forward Synthesis of Textures and Stylized Images. In *Proceedings of the 33rd ICML - Vol. 48*. 1349–1357.
- Zdravko Velinov, Marios Pappas, Derek Bradley, Paulo Gotardo, Parsa Mirdehghan, Steve Marschner, Jan Novák, and Thabo Beeler. 2018. Appearance capture and modeling of human teeth. In *SIGGRAPH Asia 2018*. 1–13.
- Steffen Wiewel, Moritz Becher, and Nils Thuerey. 2019. Latent Space Physics: Towards Learning the Temporal Evolution of Fluid Flow. In *CGF*, Vol. 38. 71–82.
- Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 2015. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on CVPR*. 1912–1920.
- You Xie, Erik Franz, Mengyu Chu, and Nils Thuerey. 2018. tempoGAN: A Temporally Coherent, Volumetric GAN for Super-resolution Fluid Flow. *ACM ToG* 37, 4 (2018).
- Xinchen Yan, Jimei Yang, Ersin Yumer, Yijie Guo, and Honglak Lee. 2016. Perspective Transformer Nets: Learning Single-view 3D Object Reconstruction Without 3D Supervision. In *Proceedings of the 30th International Conference on NIPS*. 1704–1712.
- Cheng Yang, Xubo Yang, and Xiangyun Xiao. 2016. Data-driven projection method in fluid simulation. *Computer Animation and Virtual Worlds* 27, 3-4 (2016), 415–424.