

A Multi-Pass GAN for Fluid Flow Super-Resolution

Maximilian Werhahn
Technical University of Munich
maxi.werhahn@gmx.de

Mengyu Chu
Technical University of Munich
mengyu.chu@tum.de

You Xie
Technical University of Munich
you.xie@tum.de

Nils Thuerey
Technical University of Munich
nils.thuerey@tum.de

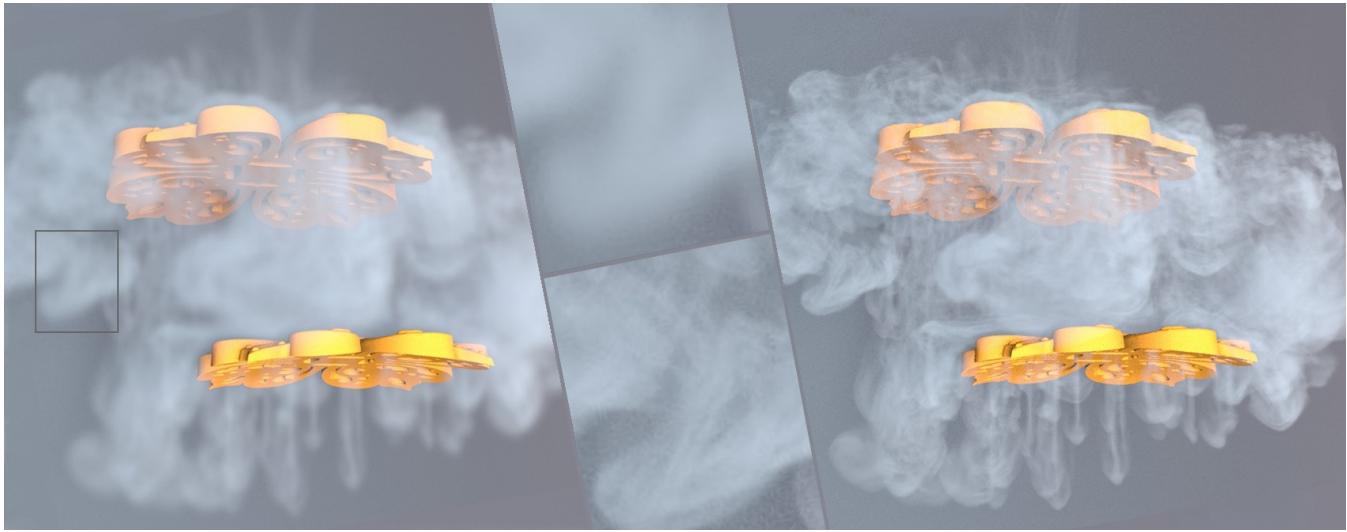


Figure 1: A 100^3 simulation (left) is up-sampled with our multi-pass GAN by a factor of 8 to a resolution of 800^3 (right). The generated volume contains more than 500 million cells for every time step of the simulation. In the middle inset, the left box is repeated as zoom-in for both resolutions.

ABSTRACT

We propose a novel method to up-sample volumetric functions with generative neural networks using several orthogonal passes. Our method decomposes generative problems on Cartesian field functions into multiple smaller sub-problems that can be learned more efficiently. Specifically, we utilize two separate generative adversarial networks: the first one up-scales slices which are parallel to the XY-plane, whereas the second one refines the whole volume along the Z-axis working on slices in the YZ-plane. In this way, we obtain full coverage for the 3D target function and can leverage spatio-temporal supervision with a set of discriminators. Additionally, we demonstrate that our method can be combined with curriculum learning and progressive growing approaches. We arrive at a first method that can up-sample volumes by a factor of eight along each dimension, i.e., increasing the number of degrees of freedom by 512. Large volumetric up-scaling factors such as this one have previously not been attainable as the required number of weights in the neural networks renders adversarial training runs

prohibitively difficult. We demonstrate the generality of our trained networks with a series of comparisons to previous work, a variety of complex 3D results, and an analysis of the resulting performance.

CCS CONCEPTS

- Computing methodologies → Neural networks; Physical simulation.

KEYWORDS

physics-based deep learning, generative models, computer animation, fluid simulation

ACM Reference Format:

Maximilian Werhahn, You Xie, Mengyu Chu, and Nils Thuerey. 2019. A Multi-Pass GAN for Fluid Flow Super-Resolution. In *Proceedings of SCA: Symposium on Computer Animation (SCA 2019)*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Deep learning-based generative models, in particular generative adversarial networks (GANs) [11], are widely used for synthesis-related learning tasks. GANs contain a generator, which can be trained to achieve a specific task, and a discriminator network that efficiently represents a learned loss function. The discriminator drives the generated data distribution to be close to the reference

data distribution. There is no explicit content-based loss function for training the generator, and as a consequence, GANs perform very well for problems with multiple solutions, i.e., multimodal problems, such as super-resolution (SR) tasks. Most of the generative models focus on two-dimensional (2D) data, such as images, as processing higher dimensional data quickly becomes prohibitively expensive. However, as our world is full of three-dimensional (3D) objects, learning 3D relationships is an important and challenging task. In the following, we will propose a volumetric training pipeline called *multi-pass GAN*. Our method breaks down the generative task to infer large Cartesian density functions in space and time into multiple orthogonal passes, each of which represents a much more manageable inference problem.

We target a fluid flow setting where the multi-pass GAN solely uses 2D slices for training but is able to learn 3D relationships by processing the data multiple times. This is especially beneficial for training: finding a stable minimum for the coupled non-linear optimization of a GAN training grows in complexity with the number of variables to train. Our multi-pass network significantly reduces the number of variables and in this way stabilizes the training, which in turn makes it possible to train networks with complexities that would be infeasible with traditional approaches. Our approach can potentially be used in a variety of 3D training cases, such as 3D data SR, classification, or data synthesis. In the following, we demonstrate its capabilities for fluid flow SR, more specifically for buoyant smoke SR, in conjunction with the tempoGAN [46] and the progressive growing of GANs [45] architectures.

For SR problems, there are inherent challenges. First of all, most SR algorithms focus on $4\times$ 2D data SR. The large size of 3D volumetric data makes it difficult to directly employ 2D SR algorithms. Existing 3D SR algorithms, such as tempoGAN, can only be applied for relatively small up-scaling factors, whereas larger factors become utterly expensive and difficult to train. Larger factors directly imply that the unknown function to be learned, i.e., the detail missing in the low-resolution (LR) input, contains content with higher frequencies. As such, the content is more complex to represent and it is more challenging to ensure its temporal coherence. A popular direction within the field of GANs is the progressive growing approach [45], which can achieve large up-scaling factors. However, the progressive growing of GANs has only been demonstrated for 2D content, as 3D volumetric data requires processing data and representing functions that have orders of magnitude more degrees of freedom. This is one of the key motivations for our approach: leveraging multi-pass GANs to decrease the resource consumption for training in order to arrive at robust learning of 3D functions.

Simulating fluid flows at high-resolution (HR) is an inherently challenging process: the number of underlying numerical computations typically increases super-linearly when the discretization is refined. In addition to the spatial degrees of freedom, the temporal axis likewise needs to be refined to reduce numerical errors from time integration. While methods like synthetic turbulence methods [21] typically rely on a finely resolved advection scheme, the deep learning-based tempoGAN algorithm arrived at a frame-by-frame super-resolution scheme that circumvents the increase in computational complexity by working on independent frames of volumetric densities via a spatio-temporal discriminator supervision. Similar to tempoGAN, our multi-pass GAN also aims at fluid simulation

SR, but targets two extra goals: reducing the unstable and intensive computations of the 3D training process and increasing the range of possible up-scaling factors. To achieve those targets, multi-pass GAN adopts a divide-and-conquer strategy for 3D training. Instead of training with 3D volume data directly, multi-pass GAN divides the main task, i.e., learning the 3D relationship between LR and HR, into two smaller sub-tasks by learning 2D relationships between LR and HR of two orthogonal planes. In a series of experiments, we will demonstrate that this strategy strongly reduces the required computational resources and stabilizes the coupled non-linear optimization of GANs.

To summarize, the main contributions of our work are:

- a novel multi-pass approach for robust network training with time sequences of 3D volume data,
- extending progressive training approaches for multi-pass GANs, and
- combining progressive growing with temporal discriminators.

In this way, we arrive at a first method for learning $8\times$ up-scaling of 3D fluid simulations with deep neural networks.

2 RELATED WORK

Large disparities between HR screens in household devices and LR data have driven a large number of advances in SR algorithms, which also find applications in other fields like surveillance [48] and medical image processing[17]. As deep learning-based methods were shown to generate state-of-the-art results for SR problems, we focus on this class of algorithms in the following. At first, convolutional neural networks (CNNs) were used in SRCNN [9], and were shown to achieve better performance than traditional SR methods [41]. Theoretically, deeper CNNs can solve more arduous tasks, but overly deep CNNs typically cause vanishing gradient problems. Here, CNNs with skip connections, such as residual networks [12], were applied to alleviate this problem[25]. Instead of generating images directly, VDSR [20] and ProSR [45] apply the CNNs to learn the residual content, which largely decreases the workload for the networks and improves the quality of the results. In our structure, we also applied residual learning to improve performance and reduce resource consumption.

Beyond architectures, choosing the right loss function is similarly crucial. Pixel-wise differences are the most basic ones and are widely-used, e.g., in DRRN [38], LapSRN [23] and SRDensseNet [43]. These methods achieve reasonable results and improve the performance in terms of the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM). However, minimizing vector norm differences typically blurs out small scale details and features, which leads to high perceptual errors. With the help of a pre-trained VGG network, perceptual quality can be improved by minimizing differences of feature maps and styles between generated results and references [14]. However, VGG is trained with large amounts of labeled 2D image data, and, for fluid-related datasets, no pre-trained networks exist. Alternatively, GANs [11] can improve perceptual quality significantly. In addition to the target model as generator, a discriminator is trained to distinguish generated results from references, which guides the generator to output more realistic results. Instead of minimizing pixel-wise differences,

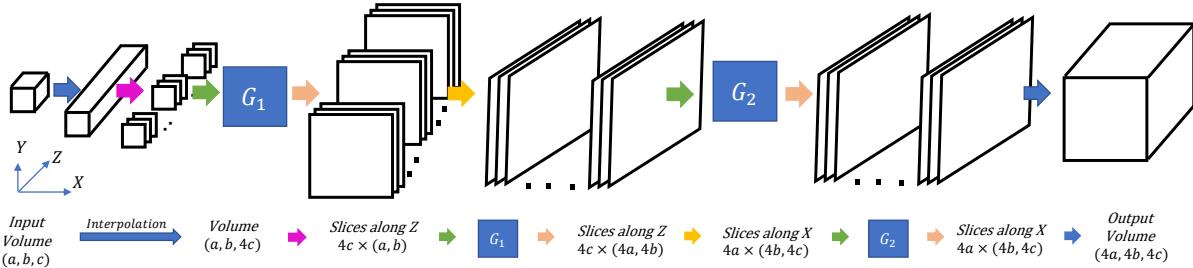


Figure 2: The pipeline of the proposed multi-pass GAN. After an up-sampling along z, we process two orthogonal directions with two adversarially trained generator networks G_1 and G_2 . The initial up-sampling ensures that all unknowns are processed evenly by the networks.

GANs aim to generate results that follow the underlying data distribution of the training data. As such, GANs outperform traditional loss functions for multimodal problems [24]. Original GANs, e.g., DC-GAN[32], adopt an unconditional structure, i.e., generators only process randomly-initialized vectors as inputs. By adding a conditional input, conditional GANs (CGAN) [28] learn the relationship between the condition and the target. Therefore, this input can be used to control the generated output. By using conditional adversarial training, EnhanceNet [35] and SRGAN [24] arrive at detailed results for natural images. tempoGAN [46] uses the velocity of the fluid simulation as a conditional input. This allows the network to learn the underlying physics and improves the resulting quality.

For SR problems, larger up-scale factors typically lead to substantially harder learning tasks and potentially more unstable training processes for recovering the details of the reference. Hence, most previous methods do not consider up-scaling factors larger than 2 or 4. However, ProSR [45] and progressive growing GANs [16] train the network in a staged manner, step-by-step. This makes it possible to generate HR results with up-scaling factors of 8 and above. Here, we also apply the progressive training pipeline to show that our method can be applied to complex 3D flow data.

The works discussed above focus on single image SR and do not take temporal coherence into account, which, however, is crucial for sequential and animation-related SR applications. One solution to keep results temporally coherent is using a sequence as input data and generating an output sequence all at once [34, 47]. This improves the temporal relationships, but frames can only be generated sequentially, and an extension to 3D data would require processing 4D volumes. For models that generate individual frames independently, an alternative solution is to use additional temporal losses. E.g., L2 losses were used to minimize differences between warped nearby frames [6, 33]. However, this typically leads to sub-optimal perceptual quality. Beyond the L2 loss, specifically designed loss terms can be used to penalize discontinuities between adjacent frames [4]. Instead of using explicit loss functions to restrict temporal relationships, temporal discriminators [8, 46] can automatically supervise w.r.t. temporal aspects. They also improve the perceptual quality of the generated sequences. In our work, a variant of such a temporal discriminator architecture is used.

To arrive at realistic fluid simulations, methods typically aim for solving the Navier-Stokes (NS) equations and inviscid Euler equations as efficiently and precisely as possible. Based on the first stable single-phase fluid simulation algorithm [37] for computer animation, a variety of extensions was proposed, e.g., more accurate advection schemes [19, 36], coupling between fluid and

solids [3, 39], and fluid detail synthesis based on turbulence models [21, 29]. In recent years, deep learning methods also attracted attention in the field of computer graphics, such as rendering [2, 5], volume illumination [15], and character control [30]. Since fluid simulations are very time-consuming, CNNs are also applied to estimate parts of numerical calculations, e.g., a CNN-based PDE solver [10, 26, 42] was proposed, as well as a fast SPH method using regression forests for velocity prediction [13]. CNNs can learn the relationships between control parameters and simulations of interactive liquids [31]. Neural networks are also used for splash generation [44] and descriptor-encoding of pre-computed fluid patches for fluid synthesis [7]. Generative networks were additionally trained to pre-compute solution spaces for smoke and liquid flows [18]. In contrast to these methods, we aim for particularly large up-scaling models that could not be realized with the aforementioned methods.

3 METHOD

Training neural networks with a large number of weights, as it is usually required for 3D problems, is inherently expensive. In an adversarial setting training large networks is particularly challenging, as stability is threatened by the size of the volumetric data. This in turn often leads to small mini-batch sizes and correspondingly unreliable gradients. In the following, we will explain how these difficulties of 3D GANs can be circumvented with our multi-pass GAN approach.

3.1 Multi-Pass GAN

To reduce the dimensionality of the learning problem, we decompose the 3D data into a stack of 2D slices and train our networks on those slices using convolutions along two spatial dimensions. As shown in Fig. 3, there are three different ways to split a 3D volume into stacked slices. If we split the 3D volume along the X-axis, we can retrieve motion information in the YZ-plane for every YZ-slice, while splitting along the Y-axis yields information about the XZ-planes.

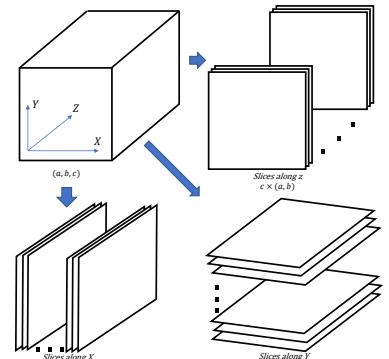


Figure 3: Three alternatives to decompose a 3D volume into stacked slices.

In order to reconstruct the target 3D function when training with 2D slices, we perform the steps illustrated in Fig. 2. First, we have to ensure that there is no bias in the inference, i.e., all spatial dimensions are processed in the same manner. Thus, for the LR volume x with resolution (a, b, c) , we linearly interpolate along the Z -axis to increase the resolution from (a, b, c) to $(a, b, 4c)$. We then split the new volume along the Z -axis to obtain $4c$ slices of size (a, b) . Here, we train a 2D network G_1 to up-scale these (a, b) slices to the target size of $(4a, 4b)$. This first generator produces HR volumes, but as it only works in the XY -plane, the results are only temporally coherent in this plane, and the model would still generate stripe-like artifacts in the XZ - or YZ -planes of the volumes. In order to allow the network to evaluate motion information along the Y - and Z -dimensions, we use a second generator network G_2 to refine the details of the volume along the Y - and Z -axes. Here, we cut the volume along the X -axis to obtain $4a$ slices of size $(4b, 4c)$. G_2 is then trained to refine those slices and drive the distribution of the output closer to the target function, i.e., an HR simulation. Note that we have found it beneficial to not change the resolution with G_2 , but instead let it process the full resolution data. In summary, combining the interpolation along Z , G_1 and G_2 jointly up-scale 3D volumes from (a, b, c) to $(4a, 4b, 4c)$ and ensure that all voxels are consistently processed by the two generator networks.

Theoretically, as soon as we have motion from two orthogonal directions, we are able to infer the full 3D motion. Thus, we train G_1 and G_2 with XY - and YZ -slices, respectively, such that they – in combination – obtain full motion information. To verify that this is sufficient, we also employed a third network G_3 to refine the XZ -plane with an additional pass through training on XZ -slices. Results from the first, second, and this optional third network can be found in Fig. 4, where each row illustrates examples of a different slicing direction. In the first row, for example, the volume is cut along the Z -axis, which G_1 operates on. When comparing the second column to the third one, it is notable that there is not much improvement for the XY -plane. However, looking at the second row which is a sample of the YZ -plane that G_2 focuses on, it is

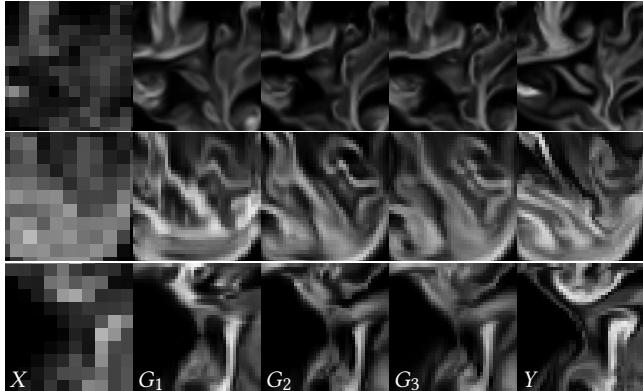


Figure 4: Comparison of the LR input density X , the output of the first network G_1 , the second network G_2 , and the third (optional) network G_3 . The HR reference Y is shown on the right. The first row depicts a slice of the XY -plane, the second row one slice of the YZ -plane, and the third row one slice of the XZ -plane. While G_2 significantly improves the output, G_3 is largely redundant.

clear that the artifacts generated by the linear interpolation are removed and the spatial coherence along the Z -axis is substantially improved. Even the slices in the XZ -plane are enhanced, as shown in the third row of Fig. 4. The fourth column in the Fig. 4 shows that G_3 leads to minimal changes along any of the axes, which matches our assumption that this additional direction of slicing is redundant. Therefore, we focus on two generator networks for our final structure, G_1, G_2 , while G_3 is not applied.

4 MULTI-PASS GAN WITH TEMPORAL COHERENCE

In general, the super-resolution problem is a multimodal task and here our goal is to additionally obtain physically plausible results with a high perceptual quality. Thus, we employ adversarial training to train both G_1 and G_2 . The full optimization problem for regular GANs can be formulated as [11]:

$$\begin{aligned} \min_G \max_D V(D, G) = & \mathbb{E}_{y \sim p_y(y)} [\log D(y)] + \mathbb{E}_{x \sim p_x(x)} [\log(1 - D(G(x)))] \\ = & \sum_m [\log D(y_m)] + \sum_n [\log(1 - D(G(x_n)))] \end{aligned} \quad (1)$$

where m, n denote the number of reference samples y and drawn inputs x , respectively. $y \sim p_y(y)$ states that reference data is sampled from the probability distribution p_y .

Our goal is similar to the one from the tempoGAN network [46]. To illustrate the advantages of our approach, we will first employ our network in a tempoGAN pipeline for $4\times$ fluid SR in this section before targeting large up-scale factors in conjunction with a progressive growing scheme in Sec. 5.

4.1 Network Structure and Loss Function

Unlike regular GANs, which only contain one discriminator, tempoGAN uses two discriminators: one spatial discriminator D_s to constrain spatial details, and one temporal discriminator D_t to keep the sequence temporally coherent. For D_t , three warped adjacent frames’ densities are used

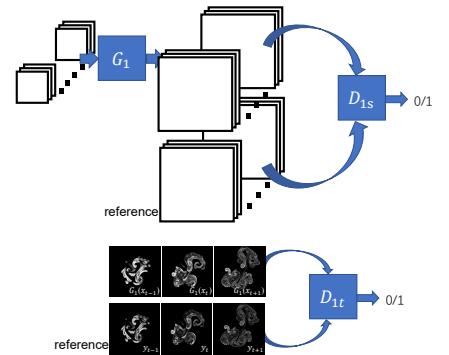


Figure 5: Spatial and temporal discriminators D_{1s} and D_{1t} supervise G_1 on the XY -plane in a multi-pass tempoGAN.

as inputs to enforce learning temporal evolution in the discriminator. The tempoGAN generator furthermore requires velocity information as input in order to distinguish different amounts of turbulent detail to be generated. Similar to tempoGAN, our generator model consists of 4 residual blocks and takes the LR density and velocity fields as inputs, which, assuming that the fields are of size 16^2 , leads to an input of size $16 \times 16 \times 4$, with one density and three velocity channels.

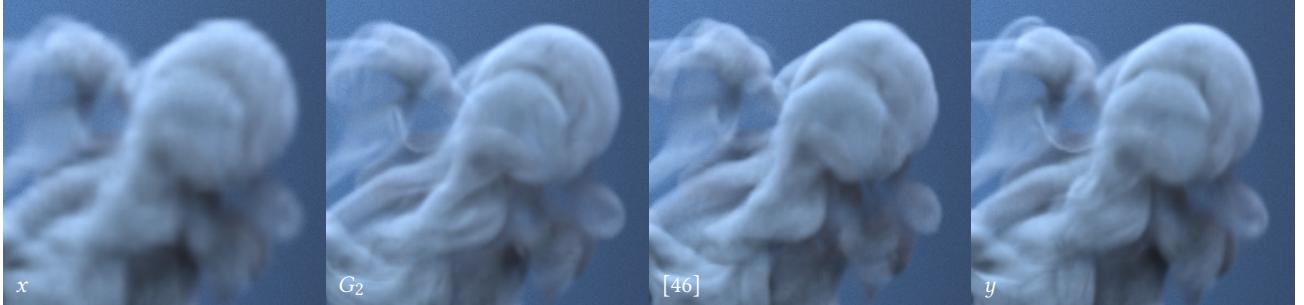


Figure 6: Comparison of LR input, output of G_2 , regular 3D tempoGAN and HR reference. The latter three have the same resolution.

For a multi-pass version of tempoGAN, there are two generators: G_1 and G_2 . Every generator is paired with a spatial and a temporal discriminator, i.e., D_{1s} , D_{1t} and D_{2s} , D_{2t} , respectively. The training process of G_1 is shown in Fig. 5. For G_2 , a similar procedure is applied, with the only difference being the up-scale operation in the network and the input data which consists of the output of the first network and LR velocities.

During the training of G_1 and G_2 , we use an additional L1 loss and feature space loss terms to stabilize the training. The resulting loss functions for training G_1 , D_{1s} , D_{1t} are then given by:

$$\begin{aligned} \mathcal{L}_{G_1}(D_{1s}, D_{1t}, G_1) = & -\sum_n [\log D_{1s}(x, G_1(x))] - \sum_n [\log D_{1t}(\tilde{G}_1 \mathcal{A}(\tilde{x}))] \\ & + \sum_{n,j} \lambda_f^j \|F^j(G_1(x)) - F^j(y)\|_2^2 + \lambda_{L_1} \sum_n \|G_1(x) - y\|_1 \end{aligned} \quad (2)$$

$$\begin{aligned} \mathcal{L}_{D_{1t}}(D_{1t}, G_1) = & -\sum_m [\log D_{1t}(\tilde{Y}_\mathcal{A})] - \sum_n [\log (1 - D_{1t}(\tilde{G}_1 \mathcal{A}(\tilde{x})))] \\ \mathcal{L}_{D_{1s}}(D_{1s}, G_1) = & -\sum_m [\log D_{1s}(x, y)] - \sum_n [\log (1 - D_{1s}(x, G_1(x)))], \end{aligned} \quad (3)$$

where \mathcal{A} advects a given frame with the current velocity, s.t. $y^t = \mathcal{A}(y^{t-1}, v_y^{t-1})$. $\tilde{G}_\mathcal{A}(\tilde{x})$ denotes three advected, consecutive generated frames: $\tilde{G}_\mathcal{A}(\tilde{x}) = \{\mathcal{A}(G(x^{t-1}), v_x^{t-1}), G(x^t), \mathcal{A}(G(x^{t+1}), -v_x^{t+1})\}$, while $\tilde{Y}_\mathcal{A}$ denotes three ground truth frames: $\tilde{Y}_\mathcal{A} = \{\mathcal{A}(y^{t-1}, v_x^{t-1}), y^t, \mathcal{A}(y^{t+1}, -v_x^{t+1})\}$. j is a layer in our discriminator network, and F^j denotes the activations of the corresponding layer. G_2 , D_{2s} , and D_{2t} are trained analogously. Note that the fluid is only advected in the XY-plane and in the YZ-plane for G_1 and G_2 , respectively.

Exemplary outputs of G_1 and G_2 from a multi-pass tempoGAN are shown in Fig. 7. While the output of G_1 still contains visible and undesirable interpolation artifacts, they are successfully removed by G_2 . Fig. 6 shows a comparison of the original tempoGAN model

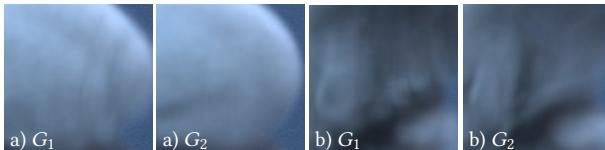


Figure 7: Example volumes generated with a multi-pass tempoGAN, shown after the first and second pass. The application of G_2 removes the stripe-like artifacts from G_1 in a) and adds new details along the yet unseen axis in b). All examples were extracted from former frames of the simulation shown in Fig. 6 to emphasize the clear differences between the passes.

and our multi-pass version at the same target resolution. Our approach yields a comparable amount of detail and visual quality, while being the result of a much simpler and faster learning problem. We will evaluate this aspect of our approach in more detail below.

5 MULTI-PASS GAN WITH GROWING

Due to the coupled non-linear optimization involving multiple networks, GANs are particularly hard to train. As this challenge grows with larger numbers of weights in the networks, techniques such as curriculum learning [45] and progressive growing [16] were proposed to alleviate these inherent difficulties. In the following, we will outline both techniques briefly before explaining how to adopt them for our multi-pass GAN in order to increase the upscaling factor to 8.

Curriculum learning describes the process of increasing the difficulty of a training target over time. In our case, we first train the generator and discriminators on pre-computed density references that are twice as large as the input. After training for a number of iterations, typically 120k, we double the up-scaling factor until the final goal of 8 is reached. The progressive growing approach

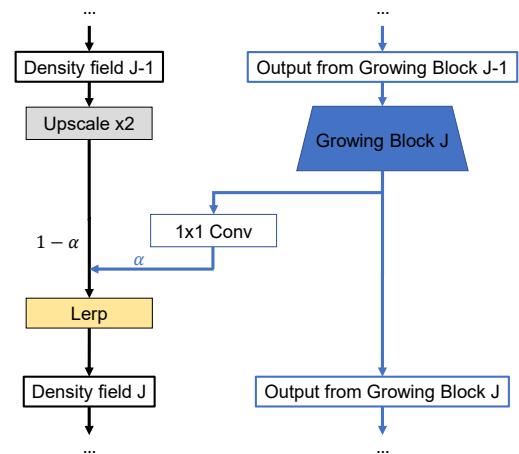


Figure 8: The structure of the progressive Growing GAN, where the parameter α is used to interpolate between the up-scaled image from the former growing block and the current one. For $\alpha = 1$, the former up-scaled density is ignored completely.

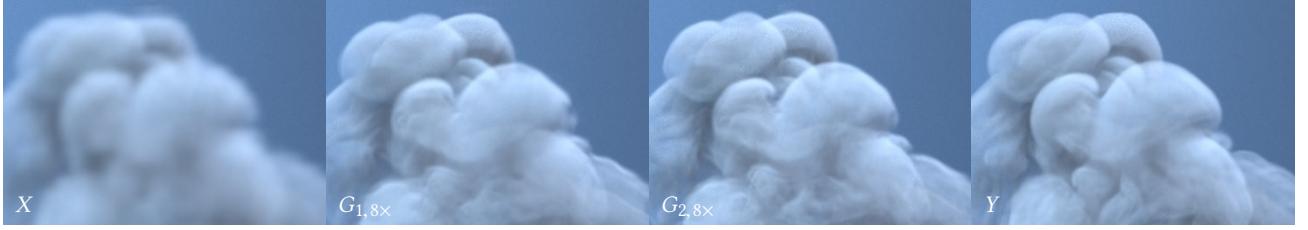


Figure 9: Comparison of LR input, output of $G_{1,8\times}$, output of $G_{2,8\times}$, and HR reference. The second generator $G_{2,8\times}$ is able to reconstruct missing smoke, sharpen smaller structures, and add novel details to the volume.

goes hand in hand with curriculum learning: after increasing the difficulty of training, additional layers are smoothly faded in for the generator and discriminator networks. This ensures that the gradients from the added stages don't impair the existing learning progress. As shown in Fig. 8, the blending process is controlled by the parameter $\alpha \in [0, 1]$, which is used to linearly interpolate between two current up-scaled density fields, e.g., combining generated outputs of scale 2 and 4. The procedure is similar regarding the discriminators. However, we use average pooling layers to down-scale the density field between the stages instead of up-scaling it. This growing technique is applied to the generator network $G_{1,8\times}$, the spatial discriminator $D_{1s,8\times}$, as well as the temporal discriminator $D_{1t,8\times}$. While fading in new layers, we increase α from 0 to 1 over the course of 120k iterations. The networks are then trained for another 120k iterations on the current SR factor. Since the task for the second generator is to purely refine the volume along the axis that was invisible to $G_{1,8\times}$, and because the only training target is the HR reference, we disable progressive growing and curriculum learning for $G_{2,8\times}$, $D_{2s,8\times}$, and $D_{2t,8\times}$.

5.1 Network Architecture

The resulting generator network $G_{1,8\times}$ consists of a sequence of ResBlocks (ResBlocks). Overall, for a total SR factor of 8, eight ResBlocks are used which are divided into four growing blocks. Each growing block up-scales the previous layer by a factor of 2, except for the first one which maintains the input resolution. For each growing block, a 1×1 convolution is trained in addition to translate the intermediate outputs into the density field if the additional down-stream layers are not yet active as shown in Fig. 8. For $G_{1,8\times}$, in order to save computations, we concentrate most of the parameters in the earlier stages of the network which have smaller spatial resolutions. We use a Wasserstein GAN loss with gradient penalty (WGAN) and a weak L1-loss in combination with the proposed equalized learning rate and pixelwise normalization (PN) [16] instead of batch normalization to keep the magnitudes in the adversarial networks under control. Thus, each ResBlock consists of the sequence of Conv3x3-ReLU-PN-Conv3x3-ReLU-PN.

In addition, inspired by VDSR [20] and ProSR [45], both our generators output residual details rather than the whole density field to improve the quality of results and to decrease computing resources. Additionally, a bi-cubic up-scaled LR density field is added to the output of $G_{1,8\times}$ and $G_{2,8\times}$. The architecture of $G_{1,8\times}$ leads to a receptive field of approximately 7 LR pixels in both dimensions. For $G_{2,8\times}$, $D_{2s,8\times}$, and $D_{2t,8\times}$, we increase the kernel size of the convolutional layers to 5 in order to increase the receptive range to

4 LR cells in each direction, resulting in a receptive field of 16 LR cells. Furthermore, we decrease the amount of feature maps to keep the number of parameters similar to $G_{1,8\times}$, $D_{1s,8\times}$, and $D_{1t,8\times}$.

Beyond our multi-pass approach, the resulting network differs from existing growing approaches in the sense that it targets functions of much larger dimensionality (3D plus time) via a temporal discriminator, while it also differs from the tempoGAN network in several ways: we employ a residual architecture with Wasserstein GAN loss and due to the progressive growing, the networks are inherently different. The exact network structure and all training parameters can be found in Appendix A and B.

Fig. 9 shows that the combination of our method and the growing approach can yield highly detailed and coherent buoyant smoke. $G_{2,8\times}$ is able to add additional details to the volume, reconstruct the desired HR shape, and generally sharpen the output along the axis that was unseen by $G_{1,8\times}$.

5.2 Loss Function

The resulting Wasserstein GAN-based loss function for our multi-pass, multi-discriminator network is then given by:

$$\begin{aligned} \mathcal{L}_{D_{1t,8\times}}(D_{1t,8\times}, G_{1,8\times}) &= -\sum_n[D_{1t,8\times}(\tilde{Y}_{\mathcal{A}})] + \sum_n[D_{1t,8\times}(\widetilde{G_{1,8\times}}_{\mathcal{A}}(\tilde{X}))] \\ &\quad + \lambda_W \sum_n[(\|\nabla_{\tilde{Y}_{\mathcal{A}}} D(\tilde{Y}_{\mathcal{A}})\|_2 - 1)^2] \\ \mathcal{L}_{D_{1s,8\times}}(D_{1s,8\times}, G_{1,8\times}) &= -\sum_m[D_{1s,8\times}(x, y)] + \sum_m[D_{1s,8\times}(x, G_{1,8\times}(x))] \\ &\quad + \lambda_W \sum_m[(\|\nabla_{\hat{y}} D(\hat{y})\|_2 - 1)^2] \\ \mathcal{L}_{G_{1,8\times}}(D_{1t,8\times}, D_{1s,8\times}, G_{1,8\times}) &= -\sum_n[D_{1t,8\times}(\widetilde{G_{1,8\times}}_{\mathcal{A}}(\tilde{X}))] - \sum_m[D_{1s,8\times}(x, G_{1,8\times}(x))] \\ &\quad + \lambda_{L1} \sum_m[\|(G_{1,8\times}(x) - y)\|_1], \end{aligned} \quad (4)$$

with $\hat{y} = (1 - R) \cdot G_{1,8\times}(x) + R \cdot y$, $\tilde{Y}_{\mathcal{A}} = (1 - R) \cdot \widetilde{G_{1,8\times}}_{\mathcal{A}}(\tilde{X}) + R \cdot \tilde{Y}_{\mathcal{A}}$, and $R \in [0, 1]$, which is randomly sampled. n and m denote the mini-batch sizes used for training the temporal and spatial discriminator and are set to 15 and 16, respectively. We use $\lambda_W = 10$ and $\lambda_{L1} = 20$.

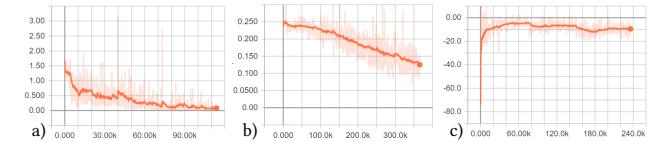


Figure 10: Comparison of the spatial discriminator training loss when using a) the full TempoGAN loss, b) the LSGAN loss, or c) the WGAN loss with gradient penalty. New layers were blended in at 60k iterations for b), c) and at 50k iterations for a).

as scaling factors for the loss terms. The networks are trained by using Adam [22] with learning rate $\eta = 0.0005$, $\beta_1 = 0.0$, $\beta_2 = 0.99$, and $\epsilon = 10^{-8}$. Eq. (4) is used instead of Eq. (2) and Eq. (3) when training $G_{1,8\times}$, $D_{1s,8\times}$, and $D_{1t,8\times}$. Note that we do not employ the feature space loss since the training process stays stable by using the WGAN-GP loss.

Training processes of the spatial discriminators with different losses are shown in Fig. 10. Here, lower values indicate that the discriminator is more successful at differentiating between real and fake images. In a) and b), which employ a tempoGAN with the loss described in Eq. (2), (3) and the least squares GAN loss [27] (LSGAN), respectively, it becomes apparent that the classification task is too easy, especially after fading in new layers. Since the losses continue to drop for tempoGAN and LSGAN which means that the spatial discriminators are growing too strong too quickly, the gradients for the discriminators will become more and more ineffective, which in turn leads to the generator receiving unreliable guidance [1]. In contrast, the stability of the WGAN with gradient penalty in c) is barely influenced by the blending process, and can recover quickly from the curriculum updates.

6 DATA GENERATION

The data used for training is generated by a stable fluids solver [37] with MacCormack advection and MiC-preconditioned CG solver in mantaflow [40]. Overall, we use 20 3D simulations with 120 frames each. For the setup, we initialize between 3 and 12 random density and velocity inflow areas combined with a randomized buoyancy force between $(0, 0, 0)$ and $(0, 3 \cdot 10^{-4}, 0)$. The LR volume is of size 64^3 , whereas the HR references are of size 256^3 for the tempoGAN $4\times$ up-scaling case, and of size 512^3 for the progressively growing GAN $8\times$ up-scaling case. The inputs x are generated by down-scaling the reference volumes y . When loading and converting the data to slices, we remove slices with average density below the threshold 0.005.. For the progressively growing version, we also generate the intermediate resolution frames of size 128^3 and 256^3 via down-scaling the references. We apply physical data augmentation methods in form of scaling and 90-degree rotations [46]. When modifying an input density field, we also have to adjust the velocity field accordingly.

While fluids in general exhibit rotational invariance, we target phenomena with buoyancy which confines the invariance to rotations around the axis of gravity. To train for shift invariance, we cut out tile pairs using randomized offsets before applying any augmentation. The size of the tiles after augmentation is 16^2 for our inputs, and $(16 \cdot j)^2$ for the targets of the current stage with j being the up-scaling factor. Overall, data augmentation is crucial for successful training runs, as illustrated with an example in Fig. 11. More varied training data encourages the generation of more coherent densities with sharper details.

7 RESULTS AND DISCUSSION

In the following, we demonstrate that our method can generate realistic outputs in a variety of flow settings, and that our network generalizes to a large class of buoyant smoke motions. For all the following scenes, please also consider the accompanying video which contains the full sequences in motion.

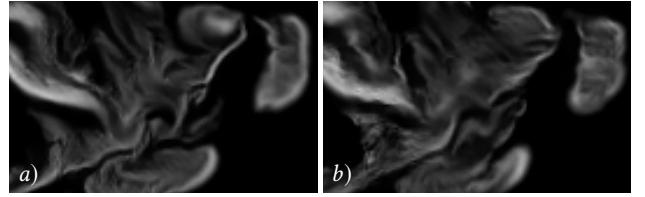


Figure 11: Comparison of a) using 90 degree rotations and flipping of the data as an additional data augmentation and b) only using scaling.

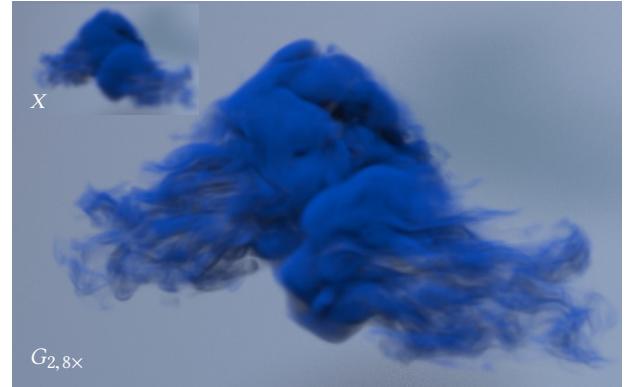


Figure 12: Compared to the training data, this colliding smoke scene contains very different motions. Based on an input of size 50^3 , the network generates large amounts of realistic detail.

In Fig. 12, a scene with three colliding smoke plumes is shown. The input of size 50^3 is transformed into an output of 400^3 by our network. The smooth streaks in the input are successfully sharpened and refined by our multi-pass GAN. In addition, the generated volumes are temporally coherent along all spatial dimensions. Fig. 13 on the other hand shows a scene of fine smoke filaments interacting with a complex obstacle in the flow. The underlying simulation is of size 100^3 , while the generated output has a resolution of 800^3 . Even though the training data did not contain any obstacles, our network manages to create realistic and detailed volumetric smoke effects.

7.1 Evaluation

In addition, it is interesting to compare the performance of our models trained for different up-sampling factors with outputs at the same target resolution. To compare the $8\times$ model with the $4\times$ one, we apply the former one to a simulation of size 32^3 , which was generated by down-sampling the actual simulation of size 64^3 . This way, both models generate a final volume of size 256^3 . Fig. 14 shows renderings of the outputs. Here it becomes apparent that our $8\times$ network manages to create even sharper and more pronounced details than the $4\times$ version, despite starting from a coarser input. For this simulation, we used the time step $\Delta t_o = 0.25$ instead of $\Delta t_t = 0.5$ as for training our networks, with Δt_t and Δt_o denoting the training and output time step, respectively. Most importantly, the range of velocities of the input simulation should not exceed the range the network was trained on, otherwise artifacts occasionally

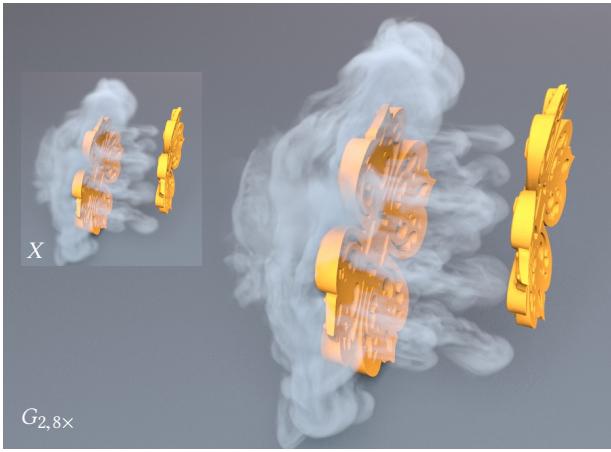


Figure 13: Even when applied to fluid-obstacle simulations, our algorithm reproduces the thin structures around objects. The resulting volume is of size 800^3 . This simulation is the same as shown in Fig. 1. Note that the network was only trained on buoyant smoke data.

occur. One possibility is to re-scale the velocities to match the training time step by multiplying with $\frac{\Delta t_t}{\Delta t_o}$ which we applied in the former example.

As shown in Fig. 15, applying the $4\times$ networks to a smooth LR simulation of size 80^3 sharpens only very few areas, e.g., the area of the plume which faces the direction of the buoyancy force. As desired, other parts such as the density inflow area are largely unchanged. Notably, however, are the few new density gradients at certain boundaries of the smoke which might be undesirable when targeting a very smooth scenarios. Since our training data did not contain such smooth simulations, the results shown here could be improved by fine-tuning the networks further.

As a simpler variant of our approach, we also evaluated applying the same network to all slices of a volume along all three dimensions, each of which was linearly up-sampled. This yields three volumes at full resolution which can be combined to obtain a single final output volume. For the combination, we have tested an averaging operation (AVG), taking the maximum per cell (MAX), or taking one of the volumes as a basis and then adding details by taking the difference between the linearly up-sampled input and the output along another axis (RES). The latter variant represents a residual transfer. Here, we also tested a variant that only used additive details to be transferred, i.e., the residuals were clamped at zero (CRES).

As shown in Fig. 16, all of these simpler methods fail at producing sharp, round edges and typically generate staircasing artifacts. The (RES) version contains especially strong artifacts while the (CRES) version does not perform much better. In comparison, our approach yields smooth and detailed outputs. Thus, training specialized networks for multiple passes is preferable over re-using a single network. The additional work to train G_2 for a second refinement pass pays off in terms of quality of the generated structures.

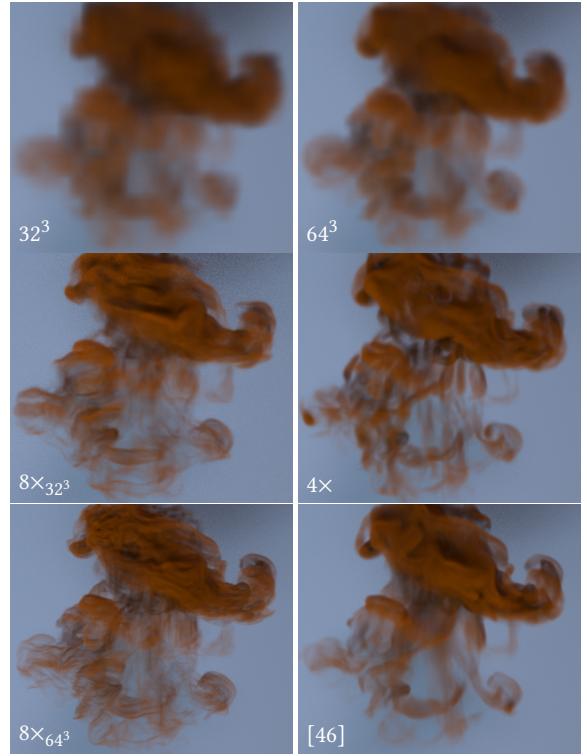


Figure 14: Comparison of $4\times$ multi-pass GAN, the $8\times$ progressively growing version and the 3D tempoGAN, all at the same final resolution. Before applying the $8\times$ model, the LR input was down-scaled once more for $8\times_{32^3}$. Originally, the simulation was of size 64^3 . In addition, the lower left image shows the result of the $8\times$ model for a 64^3 input. The generated volume contains even sharper structures.

7.2 Performance

Despite making it possible to achieve large up-scaling factors, our method also reduces training time. Training our $4\times$ multi-pass network took approximately 3 days for the first generator and 2 days for the second. This is almost twice as fast as training a 3D model reported by previous work [46]. We additionally only employed a single GTX 1080 Ti instead of two GPUs. Training the progressively growing network for an $8\times$ up-scaling took about 8 and 5 days for the first and second generator network, respectively. As this scale is not feasible with previous work, we cannot compare training times for the $8\times$ case.

Similar to previous work, the memory available in current GPUs can be a bottleneck when applying the trained networks to new input, and can make it necessary to subdivide the inputs into tiles. Regarding our implementation, we are able to deal with full slices of up to approximately 256^3 . Therefore, we usually do not need to apply tiling. For larger volumes, the tiling process requires an additional overlap for the tiles (4 LR cells for G_1 and 16 HR cells for G_2). It takes around 10.14 seconds to apply our $4\times$ network to a single frame of size 64^3 to up-scale the volume to 256^3 . Applying $G_{1,8\times}$ to a volume of size 64^3 takes 7.58 seconds while refining it with $G_{2,8\times}$ takes an additional 52.07 seconds. We compared our

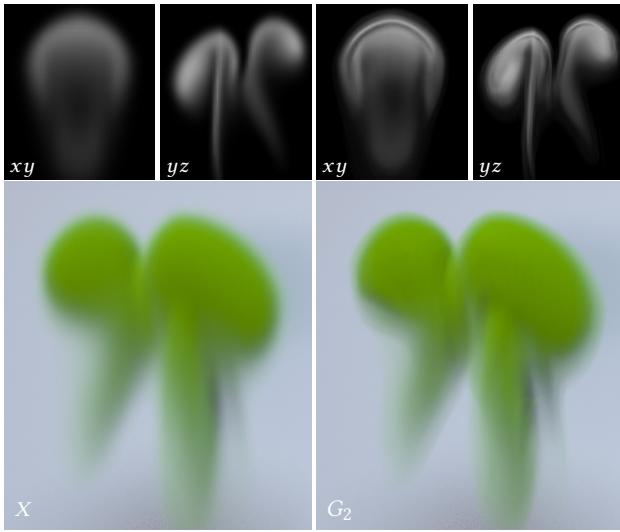


Figure 15: Applying the algorithm to a smooth input simulation with velocities of low magnitude results in the generation of barely any additional details. This is the desired outcome: the networks preserves the rather smooth regions of the input.

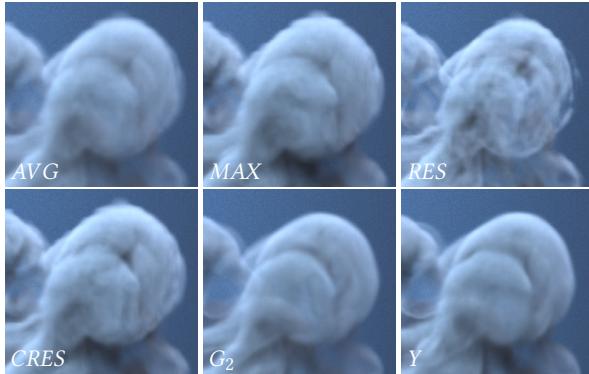


Figure 16: Different combination techniques of multiple volumes which were up-sampled along different directions. In comparison to this, our method produces sharper edges and no stripe artifacts in any dimension.

multi-pass GAN with a regular CPU-based solver using CG or multi-grid pressure solvers which are shown in Table 1. According to the CFL condition, we applied a $\frac{1}{f}$ smaller time step (where the up-scaling factor f is 4 or 8 in our case) for stability and convergence. E.g., if we generate data of the same time length as with the 8× multi-pass GAN, we run 8× more iterations with a regular or multi-grid fluid solver with a $\frac{1}{8}$ time step. Note however, that this solver is CPU-based and therefore it is difficult to compare the actual timings. Applying our method scales linearly with the number of cells of the simulation, whereas fluid solvers with CG-based pressure projections typically scale super-linearly. Besides, for a regular solver, simulation time noticeably rises when the smoke volume increases in later frames while the content does not affect generation time for the multi-pass GAN. In Table 1, we clearly see

Table 1: Evaluation of Performance (avg. for 100 frames)

	Regular solver		Multi-grid solver		Multi-pass GAN	
Resolution	256	512	256	512	256	512
Computation time (s)	116.90	1376.54	41.90	463.81	10.14	59.65

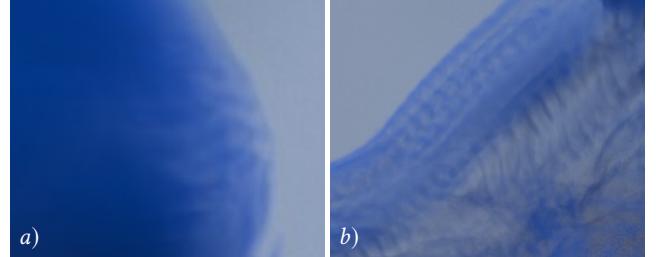


Figure 17: Examples of patterns that our 8× network generates in unfavorable situations.

that the multi-pass GAN is significantly faster than regular and multi-grid fluid solvers when generating data for a given number of frames.

7.3 Limitations and Outlook

A first limitation of our approach is that it requires multiple passes over the full HR volume (two in our case). However, in practice, these passes often result in speed-ups compared to networks that have to process the full data set at once. E.g., the tempoGAN architecture relies on up to 128 latent features per volumetric degree of freedom. These intermediate representations have to be stored on the GPU and quickly fill up the memory capacities of current GPU architectures. Hence, larger output volumes will require tiling of the inputs and induces increased workloads due to these ghost layers. In contrast, our multi-pass networks only build feature spaces for slices of reduced dimensionality that typically can be processed without tiling.

Our 8× network in its current form can lead to artifacts caused by the initial LR sampling. This typically only happens in rare situations, two examples of which are shown in Fig. 17. Here, the network seems to reinforce steps from the input data. These most likely stem from the difficult inference task to generate a 512× larger output. As we have not observed these artifacts in our 4× versions, they potentially could be alleviated by larger and more varied training data sets.

8 CONCLUSION

We have presented a first multi-pass GAN method to achieve volumetric generative networks that can up-sample 3D spatio-temporal smoke effects by factors of eight. We have demonstrated this for the tempoGAN setting and demonstrated that our approach can be extended to curriculum learning and growing GANs. Most importantly, our method reduces the number of weights that need to be trained at once, which yields shorter and more robust training runs. While we have focused on single-phase flow effects in 3D, our method is potentially applicable to all kinds of high-dimensional

field data, e.g., it will be highly interesting to apply our method to 4D space-time data sets.

9 ACKNOWLEDGMENTS

This work was funded by the ERC Starting Grant realFlow (StG-2015- 637014). We would like to thank Marie-Lena Eckert for paper proofreading.

APPENDIX

A NETWORK ARCHITECTURE

Table 2: Architecture of $G_{1,8\times}$

Layer	Output Shape
LR Input	16×16×4
ResBlock. 3×3	16×16×16
ResBlock. 3×3	16×16×64
Avg-Depool	32×32×64
ResBlock. 3×3	32×32×128
ResBlock. 3×3	32×32×64
Avg-Depool	64×64×64
ResBlock. 3×3	64×64×64
ResBlock. 3×3	64×64×32
Avg-Depool	128×128×32
ResBlock. 3×3	128×128×32
ResBlock. 3×3	128×128×16
Conv. 1×1	128×128×1
total parameters	550k

Table 3: Architecture of $D_{1s,8\times}$ & $D_{1t,8\times}$

Layer	Output Shape
HR Input	128×128×{2, 3}
Conv. 1×1	128×128×32
Conv. 3×3	128×128×32
Conv. 3×3	128×128×64
Avg-Pool	64×64×64
Conv. 3×3	64×64×64
Conv. 3×3	64×64×128
Avg-Pool	32×32×128
Conv. 3×3	32×32×128
Conv. 3×3	32×32×128
Avg-Pool	16×16×128
Conv. 3×3	16×16×32
Conv. 3×3	16×16×4
Flatten & FC	1×1×1
total parameters	470k

Table 4: Architecture of $G_{2,8\times}$

Layer	Output Shape
LR Input	16×16× k_i
ResBlock. 5×5	64×64×12
ResBlock. 5×5	64×64×48
ResBlock. 5×5	64×64×96
ResBlock. 5×5	64×64×48
ResBlock. 5×5	64×64×48
ResBlock. 5×5	64×64×24
ResBlock. 5×5	64×64×24
Conv. 1×1	64×64×1
total parameters	774k

Table 5: Architecture of $D_{2s,8\times}$ & $D_{2t,8\times}$

Layer	Output Shape
HR Input	64×64×{2, 3}
Conv. 1×1	64×64×24
Conv. 5×5	64×64×24
Conv. 5×5	64×64×48
Conv. 5×5	64×64×48
Conv. 5×5	64×64×96
Conv. 5×5	64×64×96
Conv. 5×5	64×64×96
Conv. 5×5	64×64×32
Conv. 5×5	64×64×4
Flatten & FC	1×1×1
total parameters	773k

Architectures of $G_{1,8\times}$, $G_{2,8\times}$, $D_{s,8\times}$, and $D_{t,8\times}$ are listed in Table 2, Table 3, Table 4, and Table 5, respectively. ReLU is used for the generator and an additional pixel-wise normalization layer is added after every convolutional layer. Each convolutional layer in the discriminator uses leaky ReLU as an activation function. The input to the spatial discriminator consists of the up-scaled LR and HR density fields, whereas the input to the temporal one is composed of three advected frames. Input to $G_{1,8\times}$ are the LR density and velocity, whereas the input to $G_{2,8\times}$ consists of LR density, velocity, and the output of $G_{1,8\times}$. Note that the tile sizes for $G_{2,8\times}$ and $G_{1,8\times}$ are 64^2 HR and 16^2 LR pixels, respectively.

B TRAINING PARAMETERS

For an up-scaling factor of 8, 3 stages exist in which new layers are faded in. 'blend iter.' describes the number of training iterations for $G_{1,8\times}$. The blending and stabilizing processes are applied in an alternating fashion: 120k iterations for fading in, 120k iterations for stabilizing, etc. This leads to $120k \cdot 6 = 720k$ iterations overall. After finishing the progressive growing of the networks, we slowly decay the learning rate for 'decay iter.'

Table 6: Parameters for progressive growing

Networks	blend iter.	decay iter.	training parameters	time	batch & tile size
$G_{1,8\times}, D_{1s,8\times}$ & $D_{1t,8\times}$	120k	160k	Adam: $[\eta = 0.0005, \beta_1 = 0.0, \beta_2 = 0.99, \epsilon = 10^{-8}], \lambda_1 = 20, \lambda_W = 10$	8 days	$16, 16^2$
$G_{2,8\times}, D_{2s,8\times}$ & $D_{2t,8\times}$	-	600k	Adam: $[\eta = 0.0005, \beta_1 = 0.0, \beta_2 = 0.99, \epsilon = 10^{-8}], \lambda_1 = 20, \lambda_W = 10$	5 days	$16, 64^2$

REFERENCES

- [1] Martin Arjovsky and Léon Bottou. 2017. Towards Principled Methods for Training Generative Adversarial Networks. *CoRR* abs/1701.04862 (2017). arXiv:1701.04862 <http://arxiv.org/abs/1701.04862>
- [2] Steve Bakó, Thijs Vogels, Brian McWilliams, Mark Meyer, Jan Novák, Alex Harvill, Pradeep Sen, Tony Deroose, and Fabrice Rousselot. 2017. Kernel-predicting convolutional networks for denoising Monte Carlo renderings. *ACM Trans. Graph.* 36, 4 (2017), 97–1.
- [3] Christopher Batty, Florence Bertails, and Robert Bridson. 2007. A fast variational framework for accurate solid-fluid coupling. In *ACM Transactions on Graphics (TOG)*. ACM, 100.
- [4] Prateep Bhattacharjee and Suhkendu Das. 2017. Temporal coherency based criteria for predicting video frames using deep multi-stage generative adversarial networks. In *Advances in Neural Information Processing Systems*. 4268–4277.
- [5] Chakravarty R Alla Chaitanya, Anton S Kaplyanyan, Christoph Schied, Marco Salvi, Aaron Lefohn, Derek Nowrouzezahrai, and Timo Aila. 2017. Interactive reconstruction of Monte Carlo image sequences using a recurrent denoising autoencoder. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 98.
- [6] Dongdong Chen, Jing Liao, Lu Yuan, Nenghai Yu, and Gang Hua. 2017. Coherent online video style transfer. In *Proceedings of the IEEE International Conference on Computer Vision*. 1105–1114.
- [7] Mengyu Chu and Nils Thuerey. 2017. Data-driven synthesis of smoke flows with CNN-based feature descriptors. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 69.
- [8] Mengyu Chu, You Xie, Laura Leal-Taixé, and Nils Thuerey. 2018. Temporally Coherent GANs for Video Super-Resolution (TecoGAN). *arXiv preprint arXiv:1811.09393* (2018).
- [9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2014. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*. Springer, 184–199.
- [10] Amir Barati Farimani, Joseph Gomes, and Vijay S Pande. 2017. Deep learning the physics of transport phenomena. *arXiv preprint arXiv:1709.02432* (2017).
- [11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.
- [12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Identity mappings in deep residual networks. In *European conference on computer vision*. Springer, 630–645.
- [13] SoHyeon Jeong, Barbara Solenthaler, Marc Pollefeys, Markus Gross, et al. 2015. Data-driven fluid simulations using regression forests. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 199.
- [14] 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*. Springer, 694–711.
- [15] Simon Kallweit, Thomas Müller, Brian McWilliams, Markus Gross, and Jan Novák. 2017. Deep scattering: Rendering atmospheric clouds with radiance-predicting neural networks. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 231.
- [16] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2017. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196* (2017).

- [17] John A Kennedy, Ora Israel, Alex Frenkel, Rachel Bar-Shalom, and Haim Azhari. 2006. Super-resolution in PET imaging. *IEEE transactions on medical imaging* 25, 2 (2006), 137–147.
- [18] Byungssoo Kim, Vinicius C Azevedo, Nils Thuerey, Theodore Kim, Markus Gross, and Barbara Solenthaler. 2018. Deep Fluids: A Generative Network for Parameterized Fluid Simulations. *arXiv preprint arXiv:1806.02071* (2018).
- [19] ByungMoon Kim, Yingjie Liu, Ignacio Llamas, and Jaroslaw R Rossignac. 2005. *Flowfixer: Using bfcc for fluid simulation*. Technical Report. Georgia Institute of Technology.
- [20] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. 2016. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1646–1654.
- [21] Theodore Kim, Nils Thürey, Doug James, and Markus Gross. 2008. Wavelet turbulence for fluid simulation. In *ACM Transactions on Graphics (TOG)*, Vol. 27. ACM, 50.
- [22] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014). arXiv:1412.6980 <http://arxiv.org/abs/1412.6980>
- [23] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. 2017. Deep laplacian pyramid networks for fast and accurate superresolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 2. 5.
- [24] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint* (2017).
- [25] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. 2017. Enhanced deep residual networks for single image super-resolution. In *CVPR*, Vol. 1. 3.
- [26] Zichao Long, Yiping Lu, Xianzhong Ma, and Bin Dong. 2017. Pde-net: Learning pdes from data. *arXiv preprint arXiv:1710.09668* (2017).
- [27] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, and Zhen Wang. 2016. Multi-class Generative Adversarial Networks with the L2 Loss Function. *CoRR* abs/1611.04076 (2016). arXiv:1611.04076 <http://arxiv.org/abs/1611.04076>
- [28] Mehdi Mirza and Simon Osindero. 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* (2014).
- [29] Rahul Narain, Jason Sewall, Mark Carlson, and Ming C Lin. 2008. Fast animation of turbulence using energy transport and procedural synthesis. In *ACM Transactions on Graphics (TOG)*, Vol. 27. ACM, 166.
- [30] Xue Bin Peng, Glen Berseth, KangKang Yin, and Michiel Van De Panne. 2017. Deeploco: Dynamic locomotion skills using hierarchical deep reinforcement learning. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 41.
- [31] Lukas Prantl, Boris Bonev, and Nils Thuerey. 2017. Pre-computed liquid spaces with generative neural networks and optical flow. *arXiv preprint arXiv:1704.07854* (2017).
- [32] Alec Radford, Luke Metz, and Soumith Chintala. 2016. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *Proc. ICLR* (2016).
- [33] Manuel Ruder, Alexey Dosovitskiy, and Thomas Brox. 2016. Artistic style transfer for videos. In *German Conference on Pattern Recognition*. Springer, 26–36.
- [34] Masaki Saito, Eiichi Matsumoto, and Shunta Saito. 2017. Temporal generative adversarial nets with singular value clipping. In *Proceedings of the IEEE International Conference on Computer Vision*. 2830–2839.
- [35] Mehdi Sajjadi, Bernhard Schölkopf, and Michael Hirsch. 2017. Enhancenet: Single image super-resolution through automated texture synthesis. In *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, 4501–4510.
- [36] 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.
- [37] Jos Stam. 1999. Stable Fluids.. In *Siggraph*, Vol. 99. 121–128.
- [38] Ying Tai, Jian Yang, and Xiaoming Liu. 2017. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1. 5.
- [39] Yun Teng, David IW Levin, and Theodore Kim. 2016. Eulerian solid-fluid coupling. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 200.
- [40] Nils Thuerey and Tobias Pfaff. 2018. MantaFlow. <http://mantaflow.com>.
- [41] Radu Timofte, Vincent De Smet, and Luc Van Gool. 2014. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian conference on computer vision*. Springer, 111–126.
- [42] Jonathan Tompson, Kristofer Schlachter, Pablo Sprechmann, and Ken Perlin. 2017. Accelerating eulerian fluid simulation with convolutional networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 3424–3433.
- [43] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao. 2017. Image super-resolution using dense skip connections. In *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, 4809–4817.
- [44] Kiwon Um, Xiangyu Hu, and Nils Thuerey. 2018. Liquid splash modeling with neural networks. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 171–182.
- [45] Yifan Wang, Federico Perazzi, Brian McWilliams, Alexander Sorkine-Hornung, Olga Sorkine-Hornung, and Christopher Schroers. 2018. A Fully Progressive Approach to Single-Image Super-Resolution. *CoRR* abs/1804.02900 (2018). arXiv:1804.02900 <http://arxiv.org/abs/1804.02900>
- [46] You Xie, Erik Franz, Mengyu Chu, and Nils Thuerey. 2018. tempoGAN: A Temporally Coherent, Volumetric GAN for Super-resolution Fluid Flow. *arXiv preprint arXiv:1801.09710* (2018).
- [47] Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. 2017. Seqgan: Sequence generative adversarial nets with policy gradient. In *Thirty-First AAAI Conference on Artificial Intelligence*.
- [48] Liangpei Zhang, Hongyan Zhang, Huanfeng Shen, and Pingxiang Li. 2010. A super-resolution reconstruction algorithm for surveillance images. *Signal Processing* 90, 3 (2010), 848–859.