Deep Scattering: Rendering Atmospheric Clouds with Radiance-Predicting Neural Networks

SIMON KALLWEIT, Disney Research and ETH Zürich
THOMAS MÜLLER, Disney Research and ETH Zürich
BRIAN MCWILLIAMS, Disney Research
MARKUS GROSS, Disney Research and ETH Zürich
JAN NOVÁK, Disney Research

We present a technique for efficiently synthesizing images of atmospheric clouds using a combination of Monte Carlo integration and neural networks. The intricacies of Lorenz-Mie scattering and the high albedo of cloud-forming aerosols make rendering of clouds—e.g., the characteristic silverlining and the “whiteness” of the inner body—challenging for methods based solely on Monte Carlo integration or diffusion theory. We approach the problem differently. Instead of simulating all light transport during rendering, we pre-learn the spatial and directional distribution of radiant flux from tens of cloud exemplars. To render a new scene, we sample visible points of the cloud and, for each, extract a hierarchical 3D descriptor of the cloud geometry with respect to the shading location and the light source. The descriptor is input to a deep neural network that predicts the radiance function for each shading configuration. We make the key observation that progressively feeding the hierarchical descriptor into the network enhances the network’s ability to learn faster and predict with higher accuracy while using fewer coefficients. We also employ a block design with residual connections to further improve performance. A GPU implementation of our method synthesizes images of clouds that are nearly indistinguishable from the reference solution within seconds to minutes. Our method thus represents a viable solution for applications such as cloud design and, thanks to its temporal stability, for high-quality production of animated content.

Fig. 1. We synthesize multi-scattered illumination in clouds using deep radiance-predicting neural networks (RPNN). We combine Monte Carlo integration with data-driven radiance predictions, accurately reproducing edge-darkening effects (left), silverlining (right), and the whiteness of the inner part of the cloud.
samples is expensive even with sophisticated techniques for free-path sampling [Kutz et al. 2017; Szirmay-Kalos et al. 2017] and advanced acceleration data structures [Szirmay-Kalos et al. 2011; Yue et al. 2010]. Rendering realistic clouds with path-tracing algorithms thus requires tens of hours of computation to obtain images with acceptable error; see Figure 2.

While such cost may be acceptable in some scenarios, interactive and throughput-oriented applications, such as asset design and movie production, require frames to be computed orders of magnitude faster. These typically employ approximations based on diffusion theory [Koerner et al. 2014; Stam 1995], density estimation [Elek et al. 2012], or semi-analytic solutions [Bouthors et al. 2008] heavily reducing visual quality in favor of fast rendering. We similarly trade accuracy for higher speed, but tackle the problem by synthesizing high-order scattering using neural networks.

Deep neural networks consist of compositions of simple building blocks that, when combined in the right way, achieve state-of-the-art performance in a wide variety of computational domains [LeCun et al. 2015]. Their application to transport problems is however still relatively unexplored. We demonstrate that, when designed carefully, these networks can accurately predict the distribution of radiance in clouds due to unobscured distant lighting upon arbitrarily many scattering events.

Since light propagation is a non-local phenomenon, the prediction speed hinges on the ability to concisely express the cloud geometry that is relevant to a given shading configuration. To that end, we employ a hierarchical point stencil that accurately captures the local neighborhood and compactly approximates remote regions and the overall bulk of the cloud. We observe that feeding the stencil to the network all-at-once yields poor results. Instead, we propose to input the hierarchy progressively, with each block of the network having direct access to at most one level of the hierarchy. Such design drives the network to reflect on the multi-scale nature of light transport and reduces the number of trainable parameters. This is key for performing many fast predictions needed to render the image.

We compare our method to reference path-traced images and the state of the art in grid-resolved diffusion theory. Our results are significantly closer to reference images and render as fast as the other approximate method.

2 RELATED WORK


Diffusion Theory. Certain situations permit approximating multiple scattering with the diffusion theory, in graphics pioneered by Stam [1995] who rendered heterogeneous media by solving a discretized form of the diffusion equation on a grid. Koerner et al. [2014] modulate the diffusion coefficient to increase accuracy in regions with low density and/or high albedo. Several semi-analytic solutions based on combining transport contributions from two or more monopoles were developed for fast simulation of subsurface scattering [d’Eon and Irving 2011; Donner and Jensen 2005; Frisvad et al. 2014; Jensen et al. 2001]. Unfortunately, none of the diffusion approaches handle the highly anisotropic scattering in clouds well, and the appearance stemming from low- and mid-order scattering, e.g. silverlining, is not reproduced well.

Tabulated Multiple Scattering. The anisotropy can be better preserved when using pre-computation-based solutions. Szirmay-Kalos et al. [2005] trace a number of light paths to build illumination networks and reevaluate the transport on the fly achieving real-time
frame rates. Moon et al. [2007] proposed to tabulate scattered radiance on spherical shells surrounding a directional point emitter embedded in a granular material. An analogous approach was used by Lee and O’Sullivan [2007] to handle homogeneous continuous volumes, and further adapted to heterogeneous media by Müller et al. [2016]. All these methods are either interactive, or produce high-fidelity images, but none of them achieve both concurrently.

Specialized Methods for Rendering Clouds. Fast modeling and rendering of clouds was first addressed by Gardner [1985] who used ellipsoids with spatially varying densities obtained as a superposition of sine waves. Real-time rates can be achieved on graphics hardware by rasterizing particles [Harris and Lastra 2001], metaballs [Dobashi et al. 2000], or volume slices [Riley et al. 2004] into billboard textures and projecting them onto the image plane. At best, these approaches capture only a limited amount of (anisotropic) scattering; backscattering is often completely ignored. Neglecting high-order scattering, as proposed by Nishita et al. [1996], leads to poor results (according to current standards) in optically thick clouds: see Figure 2.

Bouthors et al. [2008] acknowledge the anisotropy by employing Lorenz-Mie scattering. We also adopt Lorenz-Mie scattering with the assumption of spherical scatters; non-spherical scatterers have been studied by Frisvad et al. [2007].

Bouthors et al. [2008] further propose to precompute solutions for flat slabs that are then fitted into the cloud geometry achieving real-time frame rates. Instead of slabs, Elek et al. [2012] record photons in a regular grid where each cell models the angular distribution using the Henyey-Greenstein (HG) basis. In a follow-up [Elek et al. 2014], the authors increased the performance by combining the HG basis with the discrete-ordinates method. Nevertheless, for albedos close to 1 the approach requires either excessive numbers of iterations or multiple super-imposed low-resolution grids that tend to blur high-frequency illumination patterns.

Our goal is to accurately synthesize all-order, anisotropic scattering and accurately capture the appearance of the dense inner part and wispy boundaries. We explore a new approach based on approximating the cloud geometry by a hierarchical descriptor and predict local illumination using a deep neural network.

3 RADIATIVE TRANSFER IN CLOUDS

In this section, we discuss properties of clouds that are key for synthesizing their appearance correctly (see Figure 3). We then briefly review the RTE to allow defining the quantity to be predicted and describing our approach in the following section.
3.1 Properties of Clouds

Clouds can be primarily categorized as a cirrus (thin, wispy curls), a stratus (hazy, featureless sheets), or a cumulus (fluffy heaps). The last type is the most challenging to render due to its high-frequency boundary details and the smooth bright appearance of the inner body stemming from many interactions that light undergoes inside.

The aerosols forming a cloud consist of many small water particles—droplets or ice crystals—the placement and size of which can be well represented statistically. The radii of droplets in a cumulus cloud range from 1 to 100 microns with means typically around 5 μ [Weickmann and Kampe 1953]. The droplet-size distribution function (DSD) has a key impact on the angular distribution of scattered light, and can be well approximated by a gamma function. Because these droplets absorb little light, long chains of light-matter interactions are an integral part of the characteristic appearance of clouds (see Figure 2). Since the scatterers are comparable in size to the wavelength of visible light, the distribution of scattered light, represented by phase function \( p \), needs to be derived from Maxwell’s equations. Lorenz [1890] and Mie [1908] developed solutions for planar waves interacting with a homogeneous set of spheres. The resulting phase function exhibits a very strong diffraction peak, a wide forward-scattering lobe, a back-scattering circular ridge, and a back-scattering peak; see Figure 4 for an example and the supplementary material for details.

As shown in Figure 3, the intricate shape of the Lorenz-Mie phase function produces characteristic visual effects such as silverlining, fogbow, and glory, which are lost or synthesized inaccurately when using an isotropic or Henyey-Greenstein approximation. At the same time, however, using the high-frequency, multi-modal Lorenz-Mie distribution makes it difficult to sample the high-energy, multi-scattered transport efficiently; too difficult to render our training data within reasonable time. We follow the suggestion of Bouthors et al. [2008] to “chop” the diffraction peak and lower the cloud density according to the fraction of scattered light contained in the peak. The explicit simulation of near-perfect forward scattering—the main source of sampling difficulties—is thus removed and accounted for implicitly by reducing the optical thickness; see the supplementary material for details.

As shown in Figure 3(d), the chopped Lorenz-Mie reduces estimation variance and produces accurate results in front-lit or side-lit configurations. However, since the slight angular blurring due to the diffraction peak is replaced by continuing in the perfectly forward direction, the very prominent effect of silverlining in backlit configurations is nearly completely lost. Since the silverlining stems from low-order scattering, we opt to use the full Lorenz-Mie phase function for the first bounce and then switch to the chopped version and its corresponding optically thinner volume. This preserves the effect of silverlining and provides sufficient performance for creating a large set of ground-truth simulations to train our networks.

3.2 Radiative Transfer

As in other works, we approximate the discrete aerosol mixture by a continuous medium, parameterized by a spatially varying extinction coefficient \( \mu_t(x) \), a globally constant albedo \( \alpha \), and an axially symmetric phase function \( p(\cos \theta) \). In such a medium, the change in radiance \( L \) at point \( x \) in direction \( \omega \) is governed by the radiative transfer equation:

\[
(\omega \cdot \nabla)L(x, \omega) = -\mu_t(x)L(x, \omega) + \mu_s(x) \int_{S^2} p(\omega' \cdot \omega) L(x, \omega') d\omega',
\]

Integrating both sides of the differential RTE along \( \omega \) yields the following Fredholm integral equation of the second kind:

\[
L(x, \omega) = \int_0^\infty \exp \left( -\int_0^\infty \mu_t(x_b) \int_{S^2} p(\omega' \cdot \omega) L(x_b, \omega') d\omega' du \right) \mu_s(x_b) \int_{S^2} p(\omega' \cdot \omega) L(x_b, \omega') d\omega' du,
\]

where \( x_a = x - u \omega \) and analogously for \( x_u \). The exponential term represents the attenuation of light on a straight line between \( x \) and \( x_a \) commonly referred to as transmittance \( T(x, x_u) \).

Since clouds are non-emissive and light can only be injected from the outside, we need to define a proper boundary condition. Assuming \( x_b \) is on a hypothetical boundary \( \Omega \) that encloses the cloud, i.e. \( \forall x_b \in \Omega : \mu_t(x_b) = 0 \), we have

\[
L(x, \omega) = \int_0^b T(x, x_u) \mu_s(x_u) \int_{S^2} p(\omega' \cdot \omega) L(x_u, \omega') d\omega' du + \int T(x, x_b) L(x_b, \omega),
\]

where \( b \) is the distance to the boundary.

Equation (1) can be estimated by recursive Monte Carlo integration: we alternate sampling of distances and directions, gradually building a path towards the boundary and evaluating the radiance thereof (e.g. by extending the path beyond \( \Omega \)). Even with the latest importance-sampling techniques for building volumetric paths, rendering high-albedo clouds by path tracing is too costly for interactive applications and we use it only for constructing the training set.

4 RADIANCE-PREDICTING NEURAL NETWORKS

We propose a supervised learning approach to sidestep the costly evaluation of the recursion in Equation (1). Our goal is to efficiently synthesize the in-scattered radiance

\[
L_i(x, \omega) = \int_{S^2} p(\omega' \cdot \omega) L(x, \omega') d\omega'.
\]

To that end, we considered learning and predicting the complete \( L_i(x, \omega) \), but observed that the network struggles with predicting uncollided radiance, \( L_i(x, \omega) \); i.e. the light that arrives from light sources directly and contributes to single scattering along camera rays. We thus opt for predicting only multi-scattered transport, precisely the indirect in-scattered radiance:

\[
L_i(x, \omega) = \int_{S^2} p(\omega' \cdot \omega) (L(x, \omega') - L_d(x, \omega')) d\omega'.
\]
We propose restricting \( \omega \) to a particular neural-network architecture that we describe in Section 4.3.

### 4.2 Descriptor Construction

The main challenge of defining a good descriptor \( z \) for a given configuration \( S \) is in capturing relevant information from the density of the cloud. In theory, a neural network should be able to extract all necessary information from raw data. However, we demonstrate that the number of trainable parameters can be significantly reduced by providing the network with an easy-to-compute feature descriptor. Managing the complexity of the network in this way has the benefit of speeding up the training procedure. More importantly, however, it speeds up the predictions; millions of which are needed to synthesize a single frame.

To represent small-scale detail near \( x \), but also the overall shape of the cloud, we employ a hierarchy of points \( z \) that sample the cloud density. The stencils at individual levels are identical, but their support increases geometrically with scaling factor \( 2^{k-1} \), where \( k \) is the (1-based) index of the level; we use \( K = 10 \) levels in total. The stencil in the highest level covers a \( (2^{9})^3 \approx 10^8 \times \) larger volume than the stencil at \( k = 1 \). In order to make the descriptor scale independent, we express the stencil dimensions and density values in terms of the mean free path; we do so by scaling the cloud geometry and density such that one unit of distance corresponds to one mean free path within the average density of the cloud; the average is computed over all non-zero voxels. We refer to the scaled density as \( \rho \).

All stencils share the same local frame centered at \( x \). The \( z \)-axis of the frame points towards the light source (e.g. the Sun or another distant emitter), the \( x \)-axis is perpendicular to the plane formed by the \( z \)-axis and \( \omega \), and the \( y \)-axis is orthogonal to the two. The stencil at each level is formed by 225 points. Initially, these points form a grid in an axis-aligned box with \([-1, -1, -1]\) and \([1, 1, 3]\) being two opposing corners. The stencil at level \( k \) is then further scaled by \( 2^{k-1} \); see Figure 5 for an illustration. For each level \( k \), we extract the density values \( \Sigma = \{ \rho(q_1^{k}), \rho(q_2^{k}), \ldots, \rho(q_{225}^{k}) \} \), the union of which, \( \Sigma = \bigcup_{k=1}^{K} \Sigma^{k} \), represents the complete stencil.

The hierarchical stencil describes the cloud density with respect to location \( x \), direction towards the light source \( \omega \), and the plane that \( \omega l \) forms with \( \omega \). We also add the angle \( \gamma = \cos^{-1}(\omega : \omega l) \) to complete the description of the configuration, i.e. \( z = (\Sigma, \gamma) \).

### 4.3 Network Architecture

We use an MLP architecture with three 2-layer blocks with residual connections and an element-wise non-linear activation function to obtain the layer output. The upper part of the network consists of 2-layer blocks (blue) with residual connections, each of which processes the output of the previous block and one level of the stencil. We also add 3 standard layers (green) at the end. In all results we use \( K = 10 \), i.e. the complete network contains 23 layers.

For the the rest of the transport we employ MC integration that samples visible points of the cloud (first integral in Equation (1)) and estimates single scattering thereof.

---

**Fig. 5.** The descriptor consists of \( 5 \times 5 \times 9 \) stencils (a) that are progressively scaled and turned into a hierarchy (b), which is then oriented towards the light source (c); only two levels of the hierarchy are shown here (red and blue) to avoid clutter.

**Fig. 6.** We feed individual levels of the hierarchical stencil progressively. The upper part of the network consists of 2-layer blocks (blue) with residual connections, each of which processes the output of the previous block and one level of the stencil. We also add 3 standard layers (green) at the end. In all results we use \( K = 10 \), i.e. the complete network contains 23 layers.
We use the rectified linear unit (ReLU) activation function $V$.

The direct and indirect components are added, multiplied by $\max$, and finally generated, simulated, and artist-created clouds are used for training or validation. The remaining 75% of the clouds were used to generate training and validation data.

### 5.1 Data Generation

To compute a single training/validation record $(z_i, L_i(x, \omega_i))$, we sample a configuration $S_i$, estimate the indirect in-scattered radiance $L_i(x, \omega_i)$ thereof, and extract the descriptor $z_i$. The shading location $(x_i, \omega_i)$ is picked by first sampling a random direction $\omega_i$ uniformly over the solid sphere. Next, a ray origin $x$ is picked such that the cloud’s bounding sphere’s cross section is hit with a uniform probability density by the ray parametrized by $(x, \omega_i)$. Lastly, we trace the ray into the cloud by sampling the free-flight distance to obtain point $x$. If this step does not produce a position within the cloud, we repeat the entire process. This procedure ensures that the density of records is higher near the cloud "surface", mimicking the distribution of records as found when path tracing, as is desired for the rendering scenario described in Section 4.4.

In typical scenarios, clouds are illuminated primarily by the Sun and the environment. Since the Sun can be parameterized well by a single direction, and the illumination from the sky and the ground can be approximated as a superposition of multiple directional lights, we train only for directional light sources. We sample a random direction $\omega_i$, from which light will arrive at the shading location, and then estimate the indirect in-scattered radiance $L_i(x, \omega_i)$ using standard MC path tracing with next-event estimation; we draw samples progressively until the sample mean lies within a ±2% interval around the true answer (estimated with 95% confidence).

To study the impact of the phase function, we generated two data sets, each with $N = 15$ million samples. The first data set uses the tabulated chopped Lorenz-Mie phase function and reduced density; the version with the peak is used only at the first bounce while rendering. The second set uses a Heney-Greenstein phase function with $g = 0.857$, matching the mean cosine of the Lorenz-Mie phase function with the peak.

### 5.2 Training Procedure

We use the squared error loss function between the logarithm of the network output and the $L_i(x, \omega)$ estimated while generating data. Specifically, for a minibatch $B$ the loss is

$$\ell_B = \frac{1}{|B|} \sum_{i \in B} \left( \log (1 + g(z_i; \theta)) - \log (1 + L_i(x_i, \omega_i)) \right)^2.$$

We solve Equation (2) using stochastic gradient descent with the Adam update rule using the default learning rate, the decay parameters reported by Kingma and Ba [2014], and minibatches of size $|B| = 1000$.

Each of the two data sets described in Section 5.1 is split 75% : 25% between non-overlapping training and validation subsets. The first is used for training the parameters of the network and the second to monitor convergence and select hyperparameters.
Deep Scattering: Rendering Atmospheric Clouds with Radiance-Predicting Neural Networks

Fig. 7. Comparison of path-traced references and our approach that employs radiance-predicting neural networks (RPNN) on several different clouds illuminated by a sun-sky model. Our images accurately predict the incident illumination both due to the sun and the sky, which is estimated by sampling a number of light directions and evaluating our predictive model on them. When considering only sun-sky illumination, our method converges 24× faster than path tracing on average. Please see the supplementary material for additional results.

Fig. 8. Comparison of path-traced references, our method, and flux-limited diffusion (FLD) on three different clouds illuminated from three distinct directions. In contrast to FLD, our data-driven method correctly captures subtle contrast details (top), better preserves overall brightness of thin clouds (middle), and produces effects such as silverlining when trained on simulations of Lorenz-Mie scattering that FLD approximates by isotropic scattering. Columns on the right show zoom-ins. All FLD and RPNN images were rendered in approximately 6 seconds.
6 RESULTS

We compare the results obtained with our data-driven method to reference images rendered with a volumetric path tracer, and approximate solutions obtained with flux-limited diffusion (FLD) [Koerner et al. 2014]; we use the authors’ GPU implementation. The five clouds shown in the figures here and in the supplemental materials were excluded from the training and validation sets.

Implementation. We implemented our prediction model in TensorFlow [Abadi et al. 2015], including the camera-ray generation, free-path sampling along primary rays, and single scattering estimation to allow for an efficient integration and execution on the GPU. For evaluating the densities, we constructed a Gaussian-filtered density mipmap of the cloud to avoid aliasing artifacts and flickering in animated sequences. We used \( \sigma_i = 2^i \), where \( i \) is the mip level, as the standard deviation of the Gaussian filters. All our and FLD results were obtained using an NVIDIA Titan X (Pascal) GPU. The path-traced reference images were computed on two Intel Xeon E5-2680v3 CPUs (24 cores / 48 threads in total). We also experimented with GPU path tracing, which, however, was close to \( 2^4 \) slower than the dual-CPU setup due to the incoherent memory accesses inherent to tracing long light paths through clouds. We therefore compare our method to the faster CPU baseline. All images were rendered at \( 1280 \times 720 \) resolution. The rendering time depends linearly on the number of pixels and can be controlled by adjusting the resolution.

Comparison to Reference. Figure 7 compares our solution to a path-traced reference on a number of different clouds in varying lighting configurations. All images in the figure use a sun-sky model; see the supplementary materials for results with HDR environment maps. In most cases, our neural network predicts the radiance values accurately, producing images that are nearly indistinguishable from the reference; we recommend studying the differences using the supplemental interactive viewer. The highest errors appear in the back-lit CLOUD 4 (bottom right), where the network produces an image with slightly less contrast. This is partly due to the fact that our data set of 80 clouds consists mostly of cumulus clouds, whereas the problematic cloud is closer in shape to a stratus.

Since both methods are Monte Carlo estimators, we can quantify their efficiency in reducing noise using time to unit variance (TTUV), defined per image as the product of render time and mean pixel variance. The TTUV can also be interpreted as the variance after \( N \) seconds. Having to integrate numerically over multiple light directions reduces our speedup since much of the noise comes from the directional sampling. Hence the sun-sky lighting and the relatively low-dynamic-range envmap render generally slower; we discuss a possible remedy in Section 7. In the case of directional illumination, the speedups are lower in back-lit configurations, where most energy is contributed by single scattering that we do not accelerate.

<table>
<thead>
<tr>
<th>Scene</th>
<th>Light dir</th>
<th>Bias FLD</th>
<th>Bias RPNN (Ours)</th>
<th>Speedup RPNN (Ours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud 1</td>
<td>Side</td>
<td>3.6e-02</td>
<td>2.32e-03</td>
<td>1.21e-03</td>
</tr>
<tr>
<td></td>
<td>Front</td>
<td>2.71e-01</td>
<td>1.78e-03</td>
<td>3.54e-03</td>
</tr>
<tr>
<td></td>
<td>Back</td>
<td>1.89e-01</td>
<td>6.87e-03</td>
<td>1.57e-02</td>
</tr>
<tr>
<td>Cloud 2</td>
<td>Side</td>
<td>1.42e-02</td>
<td>1.88e-03</td>
<td>4.12e-03</td>
</tr>
<tr>
<td></td>
<td>Front</td>
<td>2.38e-02</td>
<td>1.53e-03</td>
<td>3.12e-03</td>
</tr>
<tr>
<td></td>
<td>Back</td>
<td>9.69e-02</td>
<td>5.15e-03</td>
<td>5.18e-02</td>
</tr>
<tr>
<td>Cloud 3</td>
<td>Side</td>
<td>1.28e-02</td>
<td>5.51e-03</td>
<td>3.75e-03</td>
</tr>
<tr>
<td></td>
<td>Front</td>
<td>9.82e-03</td>
<td>1.54e-03</td>
<td>1.41e-03</td>
</tr>
<tr>
<td></td>
<td>Back</td>
<td>1.49e-01</td>
<td>5.24e-03</td>
<td>1.63e-02</td>
</tr>
<tr>
<td>Cloud 4</td>
<td>Side</td>
<td>8.25e-02</td>
<td>1.51e-03</td>
<td>3.73e-03</td>
</tr>
<tr>
<td></td>
<td>Front</td>
<td>2.69e-02</td>
<td>1.75e-03</td>
<td>1.76e-03</td>
</tr>
<tr>
<td></td>
<td>Back</td>
<td>2.51e-01</td>
<td>6.35e-03</td>
<td>4.94e-02</td>
</tr>
<tr>
<td>Cloud 5</td>
<td>Side</td>
<td>1.21e-02</td>
<td>2.08e-03</td>
<td>1.51e-02</td>
</tr>
<tr>
<td></td>
<td>Front</td>
<td>2.57e-02</td>
<td>3.91e-03</td>
<td>3.16e-03</td>
</tr>
<tr>
<td></td>
<td>Back</td>
<td>4.30e-01</td>
<td>5.61e-03</td>
<td>5.18e-02</td>
</tr>
</tbody>
</table>

Comparison to Flux-limited Diffusion. In Figure 8, we compare also to FLD. Unlike our method, the diffusion theory-based FLD simulates all orders of scattering—including single scattering—with the inherent assumption of isotropic scattering. This prevents correctly synthesizing the silverlining effect, which stems from low-order, highly forward Lorenz-Mie scattering. While thin clouds are problematic for our method (due to a rather small number of those in the training set), FLD produces even worse results, overestimating the cloud brightness (middle row) and significantly changing the visual look of the cloud (see the supplementary material). In Table 1, we report the root-mean-square error on converged images of FLD and our method as bias. Our method outperforms FLD by one to two orders of magnitude in this regard.

Impact of Phase Function. In Figure 9, we compare renderings with Lorenz-Mie and Heney-Greenstein phase functions synthesized by path tracing and our method. Being an approximation, our method
with 23 layers each containing 200 nodes. The deep narrow MLP
This highlights the benefit of the progressive feeding that provides
varying resolution and a data-driven stencil with points distributed
vanilla MLP with different numbers of layers and nodes. Figure 10
7 DISCUSSION
In this section, we compare our network design to three other ar-
chitectures that we tested, analyze the impact of certain hyper-
parameters, and discuss limitations of our current approach.
Comparison to Vanilla MLP and CNN. We experimented with a
vanilla MLP with different numbers of layers and nodes. Figure 10
compares our approach to a shallow wide MLP with 5 layers con-
taining \([400, 400, 200, 200, 200]\) nodes, and to a deep narrow MLP
with 23 layers each containing 200 nodes. The deep narrow MLP
has the same number of layers as our design: we use 10 blocks, each
with 2 layers, and 3 layers at the end (each layer contains 200 nodes).
The main difference to our design is that we feed the hierarchical
descrptor to the other two variants all at once; the entire hierarchy
is input to the first layer. The other difference is the presence of
residual connections in our design. All compared MLPs use about 1.4
million trainable parameters. As demonstrated by the convergence
curves, our architecture learns faster and converges to a lower error.
This highlights the benefit of the progressive feeding that provides
means to better adapt to signals at different frequency scales.
We also tested a CNN; please see the supplementary document for
the description of the architecture. We did not see quality improve-
ments over our approach while incurring a significant performance-
and stability hit. This may be because the size of the input is rela-
tively small compared to applications where CNNs excel. Further-
more, the spatial structure which CNNs are designed for is fully
captured by the input stencil. We conjecture that if the descrip-
tor was sufficiently larger, an appropriately designed CNN might
yield a quality improvement over our network at the expense of
additional time spent constructing descriptors, which is already the
computational bottleneck of our approach.
Stencil Size. We tested several stencils and their sizes. Figure 10
compares the convergence of the validation error with stencils of
varying resolution and a data-driven stencil with points distributed
according to the mean fluence distribution around the shading lo-
cation; please see the supplementary material for details. Contrary
to our intuition, the fluence-based approach never exceeded the
performance of the straightforward regular stencils. We chose the
5 \(\times 5 \times 9\) configuration as it strikes a good balance between accuracy
and the cost of querying the density values and number of trainable
parameters in the network.
Including Direct In-scattering in Predictions. We compared the con-
vergence error of learning all in-scattered radiance \(L_s\) (green curve)
and only the indirect component \(L_i\) (blue curve). The network
struggles with learning the full \(L_s\) due to the very high-frequency nature of
direct illumination. This component is largely responsible for the
prominent silverlining effect in backlit configurations, and in-
accurate predictions would be detrimental to the overall visual qual-
ity. Since the variance of integrating the direct in-scattered radiance
via Monte Carlo is relatively low, excluding \(L_d\) from the prediction
helps attain high quality at tractable extra cost as shadow rays can
be made coherent.
Non-Cloud Shaped Objects. We tried to predict radiance on non-
cloud shaped objects using one of our models we trained on clouds
(see Figure 11 and the supplementary material). As expected, our
model performs worse than on clouds, given that the training data
is not representative of the rendered shapes. While general charac-
teristics of light transport are present, the brightness of subsurface
scattering is overestimated. These experiments highlight another
limitation of our approach: our stencil—being unable to resolve de-
tail at large distances—cannot capture distant hard shadows well.
We do not expect a proper training set to fully alleviate this problem.

![Fig. 10. Convergence of validation error for different network designs (left) and various stencil shapes with our network architecture (right).](image-url)

![Fig. 11. Comparison of path-traced references and our method on the StanfordBunny and the SphereBox scenes under directional illumination. On such shapes, which are radically different from the clouds that the network was trained on, the incurred bias is substantially higher. Furthermore, due to the coarse approximation of distant shapes by our stencil, hard shadows can not be well synthesized (bottom row). Please see the supplementary material for additional results.](image-url)
Environment Lighting. Since we estimate part of the transport using MC integration (distance sampling and single scattering), the rendered images may still suffer from a small amount of residual noise. Figure 12 compares the levels of noise with different numbers of samples per pixel (SPP). The noise is exacerbated by environmental lighting, which requires multiple samples/predictions to cover the full sphere of directions. An interesting solution to avoid sampling the environment map would be to expand it into a suitable basis, e.g., spherical harmonics, and feed the expansion coefficient into the network, thereby training it to predict the radiance due to the full sky and ground. While fruitful, we believe this topic warrants its own investigation and leave it for future work. The noise could also be removed by applying an a-posteriori denoiser.

Retraining and Adaptation. Being data-driven, the performance of our method depends largely on how well the training set represents the actual rendering scenario. That being said, our networks require ≈12h of training on a single GPU. Tailoring their performance to a particular kind of cloud or specific cloud formations is thus relatively easy, provided that enough training data is available. Lightweight retraining could also be achieved by employing the idea of progressive networks [Rusu et al. 2016].

Temporal Stability. A key strength of our approach is its temporal stability stemming from the hierarchical descriptor and filtering. The accompanying video shows sequences with an orbiting camera and light sources, a cloud with changing density, and a cloud with changing shape, all of which are temporally stable.

8 CONCLUSION

We presented a novel approach for rendering atmospheric clouds at fast rates. Our radiance-predicting neural networks represent a novel point in the spectrum of methods trading accuracy for speed. We achieve results that are nearly indistinguishable from the reference often faster or at comparable cost to previous approaches. The key and novel ingredient of our method is the progressive feeding of a hierarchical descriptor into the network. We further improve performance using standard techniques, such as residual connections and mipmap mapping. Our method is temporally stable and could present a viable solution for applications such as asset design, future computer games, and also offline rendering if a small amount of bias is visually acceptable. We demonstrated that deep neural networks can be successfully applied to transport problems and hope that our work will stimulate further developments in that direction.

ACKNOWLEDGMENTS

We thank Delio Vicini for proofreading and helpful discussions, the Stanford 3D Scanning Repository for the ARMADILLO, DRAGON, and STANFORD-BUNNY models, and Magnus Wrennne and the Pixar Animation Studios for providing most of the cloud volumes used throughout project. We are also grateful to David Koerner for sharing the source code of the flux-limited diffusion [Koerner et al. 2014].

REFERENCES


We thank Delio Vicini for proofreading and helpful discussions, the Stanford 3D Scanning Repository for the ARMADILLO, DRAGON, and STANFORD-BUNNY models, and Magnus Wrennne and the Pixar Animation Studios for providing most of the cloud volumes used throughout project. We are also grateful to David Koerner for sharing the source code of the flux-limited diffusion [Koerner et al. 2014].


2017-09-15 11:31 page 10 (pp. 1-11)
Deep Scattering: Rendering Atmospheric Clouds with Radiance-Predicting Neural Networks • 231:11

beams. ACM TOG 30, 1 (Feb. 2011), 5:1–5:19
