

# exploRNN: Understanding Recurrent Neural Networks through Visual Exploration

Alex Bäuerle, Raphael Störk, and Timo Ropinski

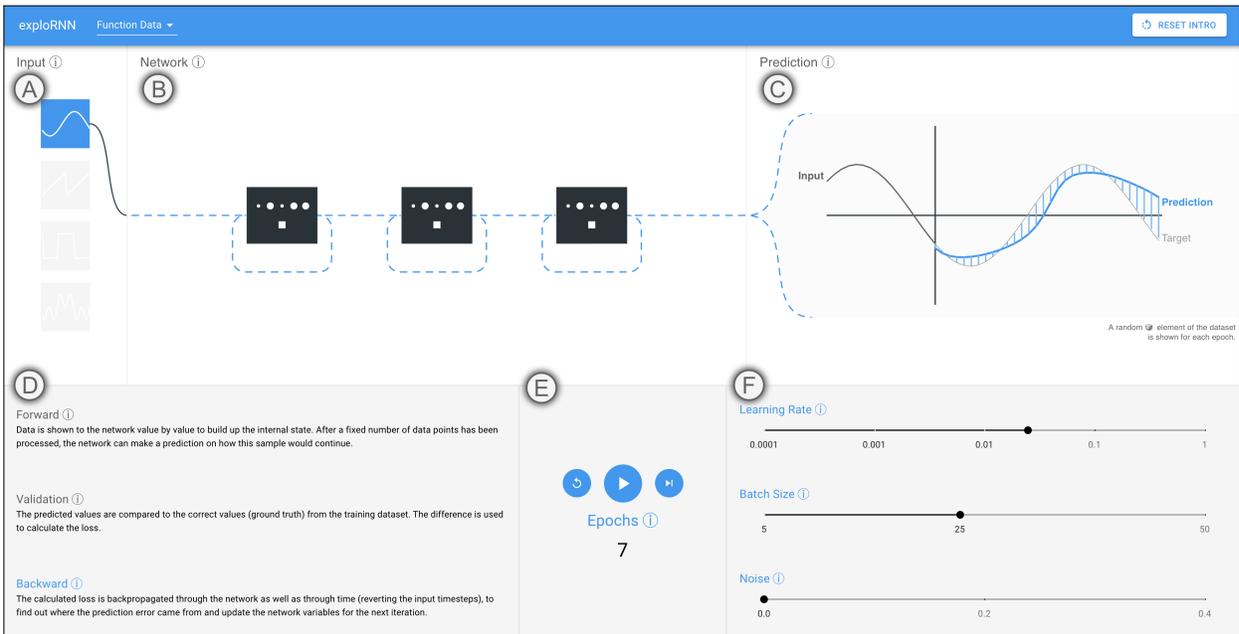


Fig. 1. Our interactive application *exploRNN* allows for exploration of recurrent neural network (RNN) architectures. (A) Simple, selectable inputs help to understand the abstract concepts behind RNNs. (B) An animated, modifiable network architecture shows the data-flow. (C) The output visualization shows the input to the network, the prediction, the ground truth, and error bars, all animated so that the user understands the temporal nature of it. (D) Textual explanations about the training process help the user understand the network. (E) Users can interactively train the RNN. (F) Modifiable parameters allow for interactive exploration of the training process.

**Abstract**— Due to the success of deep learning and its growing job market, students and researchers from many areas are getting interested in learning about deep learning technologies. Visualization has proven to be of great help during this learning process, while most current educational visualizations are targeted towards one specific architecture or use case. Unfortunately, recurrent neural networks (RNNs), which are capable of processing sequential data, are not covered yet, despite the fact that tasks on sequential data, such as text and function analysis, are at the forefront of deep learning research. Therefore, we propose *exploRNN*, the first interactively explorable, educational visualization for RNNs. *exploRNN* allows for interactive experimentation with RNNs, and provides in-depth information on their functionality and behavior during training. By defining educational objectives targeted towards understanding RNNs, and using these as guidelines throughout the visual design process, we have designed *exploRNN* to communicate the most important concepts of RNNs directly within a web browser. By means of *exploRNN*, we provide an overview of the training process of RNNs at a coarse level, while also allowing detailed inspection of the data-flow within LSTM cells. Within this paper, we motivate our design of *exploRNN*, detail its realization, and discuss the results of a user study investigating the benefits of *exploRNN*.

**Index Terms**— Explainability, deep learning, recurrent neural networks, machine learning, interactive exploration, explorable explanations

## 1 INTRODUCTION

With its recent advancements, deep learning has gained immense traction in research and industry as well as education. As more and more applications benefiting from deep learning dominate the news, and job opportunities related to machine learning are unprecedented, many want to learn about and understand these technologies. While initial progress in deep learning was mainly possible due to the rise of con-

volutional neural networks (CNNs), large training data sets, and GPU training in the context of image recognition [11, 20, 33], other network architectures, such as RNNs, which are able to process sequential data, are getting increasingly important.

As more advanced learning architectures, such as RNNs, are more difficult to comprehend, and as they introduce concepts that are fundamentally different from classical computer science, communicating how these networks are able to accomplish the desired results is important for users who want to know more about the technology. By making the process behind RNNs transparent to this target group, research in machine translation and other sequential learning tasks can be accel-

• All authors are with Ulm University. E-mail: See <https://a13x.io>

erated, as the field is opened up to additional users and contributors. Along this line, the visualization community has shown how interactive visual explorables can be effective for learning other deep learning concepts [17, 18, 23, 29].

Since the different architectures exploited bring their unique challenges, existing educational applications usually focus on one type of architecture. Unfortunately, the set of existing applications still does not cover one specific class of modern neural network types, namely RNNs. This is despite the fact, that RNNs are the tool of choice for learning on sequential data, and thus are widely adopted in tasks such as speech processing [9, 25], handwriting recognition [8], and machine translation [35], among many others. As diverse as the class of applications addressed by RNNs, are the concepts that capture temporally dependent information within RNNs. As these concepts greatly differ from other network types, we argue, that educating developers as well as consumers about RNNs could be of great benefit. In this paper, we, thus, propose *exploRNN*, an interactive explorable visualization for RNNs, that runs directly in any modern web browser. Within this paper, we make the following contributions:

- **An interactive learning environment for RNNs**, based on explorable components to make the training process of RNNs accessible to non-experts.
- **A new visualization approach for RNNs**, which allows for investigation at different levels of granularity through interactively refinable visualizations. These different levels provide the user with both, a rough overview of the functionality of RNNs, as well as detailed insight into the individual components of LSTM cells.
- **A browser-based open source implementation**, that makes broad public access to these techniques possible, and provides users with tools to learn about this important aspect of deep learning.

The design of *exploRNN* has been informed by educational objectives, which streamlined the development of our interactive visualization approach towards well-defined goals. This way, we were able to identify and communicate the most important aspects of RNNs in an educational setting. We adopt the approach of other educational visualizations in the literature [14, 15, 26], where interactive explorables are used to help learners understand abstract concepts in a fun and intuitive manner. While only the adoption after the public release can clarify the effectiveness of our interactive approach, we also conducted a first qualitative user evaluation, which sheds light on the usability and learning effect of *exploRNN*.

*exploRNN* can be experimented with online at: <https://mi-pages.informatik.uni-ulm.de/exploRNN/>, and thus contributes to a fast-growing corpus of visualization work in the field of deep learning. To our knowledge, *exploRNN* is the first educational system that is targeted towards RNNs, an important and growing class of neural networks.

## 2 BACKGROUND: RECURRENT NEURAL NETWORKS

To build a foundation for the discussions in this paper, we first provide some background knowledge on RNNs within this section. Readers well skilled in this area of learning, might skip this section for now, and return later in order to review the notation used throughout the paper.

CNNs and multi-layer perceptrons (MLPs) are currently used for most classic vision-related deep learning tasks. Here, individual data points are fed to the network, and processed layer by layer, before a prediction is made. This process only contains forward connections of data flow in order to make a prediction. For sequential data, this is, however, not sufficient, as previous inputs play an essential role in making a prediction. When for instance predicting how a text continues, looking at one character or even one word is not enough, but one needs to account for many words of characters that preceded the one to be predicted, as well as their temporal relation. For such sequential use cases, RNNs provide a cyclical architecture, in which the output of the previous timestep is used in combination with new inputs to inform the

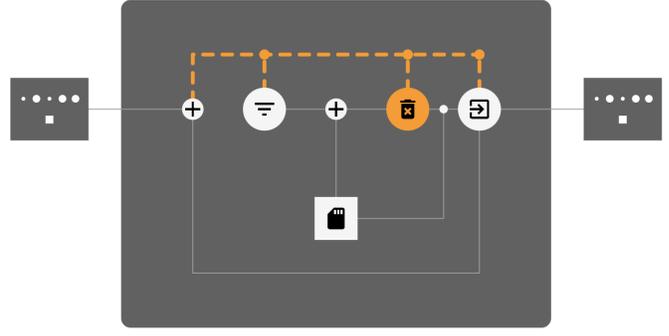


Fig. 2. LSTM cell with all its gates, inputs, and outputs. The input is added to the output of the previous time step and then used by the three gates for the gate activation. In combination with the forget gate (trashcan icon), which also takes into account the cell state (disk icon) of the previous time step, the input gate (filter icon) is used to update the cell state using this computed input/previous output combination. Finally, the output gate (arrow icon) determines based on input and cell state, what information will be output from this cell.

activation calculation of a cell. This way, RNNs are able to capture the temporal dependence of sequential data.

RNNs are trained as any other neural network, using backpropagation. The only difference is, that the prediction error is not only propagated back through the layers, but also within the recurrent connections of the layers, to be able to update the weights that handle the previous cell state. Therefore, this process is referred to as backpropagation through time (BPTT).

The most basic architecture for a RNN cell simply computes the output activation for timestep  $t$  as  $a^t = \tanh(W_{ax}x^t + W_{aa}a^{t-1} + b_a)$ , with  $x^t$  as the input at timestep  $t$  and weight matrices  $W_{ax}$  and  $W_{aa}$  and bias vector  $b_a$ . This cell architecture, however, proved to be incapable of capturing long-range dependencies in sequences due to the vanishing gradient problem [2, 12, 24], which gets more severe the further one goes back in time. This is why Hochreiter and Schmidhuber proposed long short-term memory (LSTM) cells [12]. An abstract visualization of one such cell can be seen in Fig. 2. They introduce a concept of internal memory, called the *cell state*, which is updated in a modular fashion, whereby three gating functions determine the updates to, and outputs from the cell state. The three gates are computed based on the input at timestep  $t$ ,  $x^t$  and the activation of the cell at timestep  $t - 1$ ,  $a^{t-1}$ , whereby they take on the following roles:

1. Input gate:  $i^t = \text{sigmoid}(W_{ix}x^t + W_{ia}a^{t-1} + b_i)$ , determines how much new information will be used to update the cell state  $c^t$ .
2. Forget gate:  $f^t = \text{sigmoid}(W_{fx}x^t + W_{fa}a^{t-1} + b_f)$ , determines how much of the information in the cell state  $c^t$  can be forgotten.
3. Output gate:  $o^t = \text{sigmoid}(W_{ox}x^t + W_{oa}a^{t-1} + b_o)$ , determines what part of the cell state  $c^t$  is used to compute the activation.

The cell state at timestep  $t$  is then computed as  $c^t = f^t \circ c^{t-1} + i^t \circ \tanh(W_{cx}x^t + W_{ca}a^{t-1} + b_c)$ , where  $\circ$  is the hadamard, or element-wise product, whereby the output activation is computed as  $a^t = o^t \circ \tanh(c^t)$ . Hochreiter and Schmidhuber showed that using this construct, the vanishing gradient problem can be mitigated and these cells can capture much longer ranges of dependencies [13].

While there are different architectures, that also use the concept of cell state and modular updating, such as gated recurrent units (GRUs) [6], their underlying idea does not greatly differ. However, LSTMs were the first to exploit the explained concepts, are more general in their application [34], and often outperform GRUs [3], which is why we focus on conveying networks using the LSTM architecture.

### 3 RELATED WORK

Visualization proved to be an effective way of communicating and explaining the functionality as well as the decision-making process of neural networks. Nevertheless, using visualization and exploration for educational purposes began long before the most recent deep learning era. Thus, in this section, after giving a brief overview of explorable explanation literature, we elaborate on both, the corpus of related work in the area of educational visualizations for non-experts, as well as the line of work on visualizing RNNs.

#### 3.1 Explorable Explanations

Explorable learning environments were invented long before deep learning raised awareness in the broader public. The effectiveness of this approach was investigated in the line of work by Hundhausen et. al. [14, 15]. Schweitzer and Brown then described design characteristics and evaluation of active learning settings in classrooms using visualization [26]. These early approaches show how effective visualization can be in communicating algorithmic thinking. In our approach, which aims at conveying the concept of RNNs to learners, we follow the lead of these earlier interactive teaching approaches, where explorable explanations provide great benefits for learning. Nevertheless, we are also inspired by more recent concepts, which have been proposed under the term *exploration* in the area of science education [36].

#### 3.2 Educational Neural Network Visualizations

There are numerous helpful visualizations that aim at conveying properties of neural network architectures, their functionality, or application scenarios. In one of these, users can draw the input for a CNN trained to recognize handwritten digits, and see how parts of the network get activated and interplay [10]. Similarly, *convnetjs*, a browser application to train a simple CNN shows how activations and gradients change during the training process [18].

One of the most prominent interactive educational visualization approaches for explaining how neural networks work, has been proposed by Smilkov et. al. [29]. In their explainable called *Tensorflow Playground*, one can select the properties of a neural network by means of layers and neurons per layer. They also allow the customization of certain training parameters, and deployed their application to be used directly in modern browsers. While this is a helpful visualization environment to understand the most basic concepts of neurons and how they are trained together, more advanced architectures need further, specialized visualizations to support understanding these approaches.

Similarly, in *Revacnn*, users can explore the activations of a CNN, while they can also modify the network structure and train the network in the browser [7]. *Revacnn* is designed for CNNs and thus cannot handle RNNs.

A further approach that is closely related to ours, but works on a different type of neural network is *GanLab* [17]. *GanLab* is focused on explaining the functionality of GANs, especially, how the generator and discriminator are used adversarially to yield synthetic data that resembles the data-distribution it was trained on. They also focus on showing this complex network architecture in an abstract way, and use simplistic data as training input to make the process more comprehensible. However, GANs bring their own visualization challenges, which are fundamentally different from those we found for RNNs.

There are also more application-specific visual explainables, such as the *Adversarial-Playground* proposed by Norton and Qi [23]. Here, users can experiment with and learn about adversarial examples for image-based CNNs.

As none of these visualization approaches is designed to help non-experts understand how RNNs function, with their unique concepts of memory and temporal dependence, we aim to fill this gap in the literature with *exploRNN*.

#### 3.3 RNN Visualizations

Apart from just educational visualizations, there is another line of visualization work targeted towards investigating training and prediction processes of RNNs. These approaches are mainly targeted towards

researchers who want to understand and debug their models, and thus be able to improve them.

An early approach towards visualizing RNNs was proposed by Karpathy et. al. [19]. They propose visualizing the activation of RNN cells for expert analysis. Strobel et. al. published a visualization called *LSTMVis*, in which the hidden state dynamics of RNNs are investigated [32]. They specifically demonstrate how text understanding can be analyzed through investigating the structure and change of the cell state. Here, one can see if the model learned to understand and interpret complex textual structures. They also presented *Seq2Seq-Vis*, in which sequence to sequence models can be probed to reveal errors and learned patterns [31]. Along the same line, Ming et. al. introduced *RNNVis* [22], which allows investigating networks that are used for problems in the domain of natural language processing. They analyze the functionality of individual hidden state units by observing their reaction to specific text segments.

With *RNNbow*, Cashman et. al. published a visualization, in which the gradients of RNNs can be analyzed [5]. They attribute the gradient to individual letters in a textual input sequence. This way, researchers can inspect how their models learn.

In another approach, Shen et. al. even propose visualizations for RNNs [27] operating on multi-dimensional sequence data. Here, developers can inspect hidden unit responses to get insight into the network. They visualize different models, such that experts can compare the unit response clusters for these. Like all the approaches described here, it is, however, not targeted towards explaining RNNs to laymen, which we aim for with *exploRNN*.

Insights on the effects of using exploration and visualization for learning, as well as present educational visualizations for neural networks, show how interactive exploration can help to simplify and explain modern neural network concepts. However, none of the existing explorables is designed for RNNs. On the other hand, special aspects of RNNs have seen great visualization work, which shows how important visual explanations for these architectures are. While these expert tools help to diagnose what RNNs have learned, they cannot be used by non-experts to acquire a basic understanding of the underlying techniques. We thus propose *exploRNN*, which aims at providing insight into the function of RNNs for laymen in this area of sequential learning.

### 4 EDUCATIONAL OBJECTIVES

In this section, we define the target audience as well as educational objectives that *exploRNN* has been designed for. Our approach is aimed at users who understand the fundamental concepts of deep learning, and already know about feed-forward neural networks. Without this background knowledge, an introduction to RNNs would require an introduction into feed-forward neural networks and the theory behind those first, which would extend the scope of this application, and prevent a special focus on RNNs. We let this more basic introduction to others, and therefore define the following educational objectives for *exploRNN*:

- O1 **Justification.** The first thing users should learn is, that feed-forward CNNs are not capable of handling temporal dependencies, whereas RNNs are specifically designed for this task. Users should thus learn, that RNNs use recurrence to access information about previous time steps.
- O2 **Data.** Users should then understand, that RNNs operate on sequential data, and that this data is fed into the network one data point after the other, in order to build up the cell state.
- O3 **Training.** Additionally, users should realize, that RNNs are trained in a manner very similar to CNNs, whereby the prediction error is used to update network weights. They should also comprehend that for RNNs, the error is not only backpropagated through layers, but also through time.
- O4 **LSTM Cells.** Apart from general RNN concepts, another learning objective for our users is that they understand the elements and mathematical operations that LSTM cells are built with. This includes the gates and data flow within these cells.

**O5 Task.** Finally, users should learn how these ideas can be transferred from abstract data to a more realistic problem context, such as text processing.

In general, rather than providing a purely mathematical background, or teaching learners how to implement RNNs themselves, our approach is aimed at providing learners with an intuitive understanding of RNNs. This should be a potential starting point from where to learn more about different types of RNNs, the underlying math, and how to program RNNs. Accordingly, our educational objectives do not only motivate the importance of RNNs, but are also targeted to provide insights about the input data and related tasks, as well as how the training process and LSTM cells work. We thus cover the five most important pillars for learning about RNNs. Accordingly, we see our work as a continuation of what was started by other learning environments [17, 29], which also present overviews of the presented techniques to provide learners with a general understanding rather than the full technical depth.

## 5 DESIGN CHALLENGES

Since RNN cells are a special form of neural network layers, they open up unique challenges for visualization-based education. Here, we describe the challenges that arise when aiming at the educational objectives introduced in the previous section, whereby we group the challenges into visualization design challenges, and technical challenges.

### 5.1 Visualization Design Challenges

We first discuss the visualization design challenges for an interactive learning environment targeted towards RNNs. Along this line, we have identified the following three main visualization design challenges:

**V1 Complexity.** For real-world applications, RNNs are typically trained on a large amount of complex data, that can be difficult to grasp. The same holds for network architectures, which are also often too complex to fully comprehend in their entirety. Accordingly, as we want to convey the functionality of RNNs in an interpretable and intuitive way, data and the employed network structure should be accessible. Thus, both, data and the network architecture should be simple enough to be easily understandable, but also realistic enough to make a compelling use case.

**V2 Dynamics.** A educational system for teaching RNN concepts should clarify the dynamics of the sequential data these networks operate on, as well as the dynamics of the training process. These dynamic processes have to be communicated visually, with clear separation between data type and data processing within the network. For the network dynamics, this includes visually communicating the different dynamics that forward (prediction) and backward (learning) dataflow directions bring with them.

**V3 Multiscale.** With RNNs, the different granularities to be communicated (network, cell, and cell components) bring their own challenges. These multiple scales need to be bridged and inspectable, while at the same time, the granularity at which the user currently operates must be communicated.

### 5.2 Technical Challenges

Whereas the visualization design challenges map directly to our educational objectives, the following technical challenges relate to the development of such a learning environment in an interactive and exploratory way. Accordingly, we identified the following three main technical challenges for *exploRNN*:

**T1 Training Time.** In real application scenarios, training processes can take up to several days until convergence. For an interactive application of RNNs, waiting days for convergence is obviously not feasible. Our technology thus has to be such that the network converges in minutes instead of hours or days, and thus provides instant feedback to the user. This limits the space of network designs, training data, and training objectives to be used.

**T2 Training Steps.** Normally, computation is done as fast as possible, to minimize the time needed for the network to converge. In our case, however, we want the user to be able to follow the training process, and observe individual training steps. It is thus important, to visually and temporally separate training steps and make them run so that one can visually follow this process.

**T3 Deployment.** Modern-day learning is often conducted via online courses, blog posts, or explainable web pages. While this makes such learning environments accessible to a broad audience, it also limits what can be shown to the user. Teacher intervention is not feasible in these scenarios, and the applications have to be performant and lightweight enough to allow for an interactive browser experience.

## 6 VISUALIZATION DESIGN

*exploRNN* consists of two main views, the network overview, as shown in Fig. 1, which displays the training progress at the network scale, and the cell view, shown in Fig. 3, that allows for detailed inspection of a LSTM cell.

In the network overview, following the natural reading direction of western cultures, we arrange the network from left to right, whereas on the left, one can see the input type that is currently used to train the network. Centered, we present an abstracted visualization of the network, where users can see how many layers the network contains. On the right, we further present a visualization of the prediction alongside the prediction error, to show the user how training progresses. Below these visualizations, we show information about the training process, controls for the training process, and provide means to change training parameters.

In the cell view, we show a detailed visualization of the selected cell on the left, embedded in small pictograms of neighboring cells. To the right of this cell visualization, one can see the input, target, and prediction values of the network, where new points are added as they flow through the cell. Below these visualizations, we show information about the training process, controls for the training process, and means to change training parameters, just as in the network overview.

In the following, we will describe the general concepts for the visualization design of *exploRNN*. In particular, we are going to discuss how we meet the aforementioned visualization challenges (Vx) while tackling the educational objectives (Ox) defined in Sect. 4.

### 6.1 Justification

To communicate how RNNs make predictions based on temporally dependent information (O1), we employ multiple visual guides.

In the visualization conveying the overall network structure, i.e., the network overview (Fig. 1), we employ the common visual coding for RNN layers, where a loop is added to the layer glyph. This symbolizes the feedback loop of information output at the previous timestep back to the input of a cell (V2, Dynamics). This visual encoding is further clarified in the cell view (Fig. 3). Here, we explicitly show how the input to a cell is composited of its previous output and new information for the three gating mechanisms, and explain the cell state to the user, which is important for the ability of LSTMs to capture temporal dependence (V3, Multiscale).

The temporal dependence that RNNs can capture is also visually communicated in the output plots of our approach. This can be seen in Fig. 1, and Fig. 3. Here, multiple data points are used to inform a prediction, which is visualized by a sliding gray box. This box indicates the number of data points used to build up the internal state before a prediction is made (V2, Dynamics).

### 6.2 Data

To communicate the sequential nature of RNNs (O2), we carefully crafted both, the data types that can be used for training, and the visualization of data in our approach.

Data for an explainable and interactive visualization of neural networks needs to be both, visually inspectable, and easy to understand. Therefore, both, Tensorflow Playground [29], and GanLab [17] use

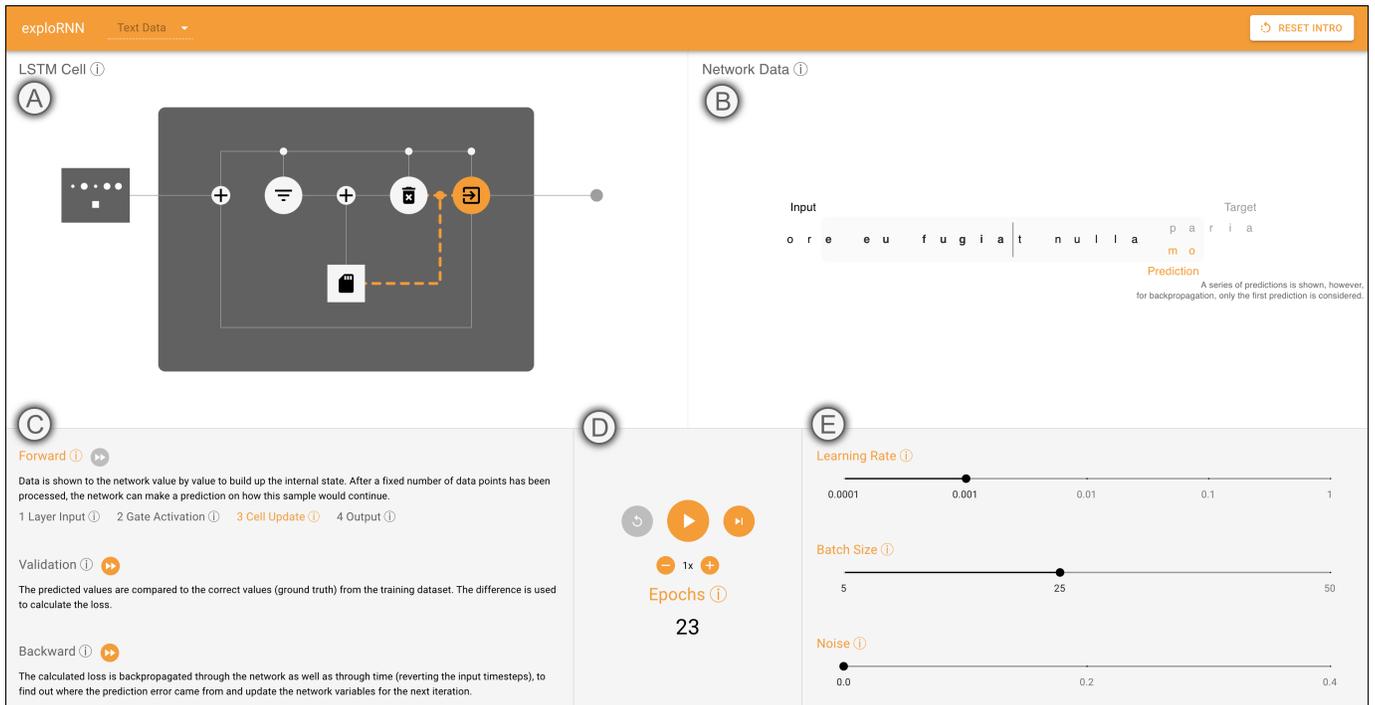


Fig. 3. In the cell view, users can inspect individual LSTM cells in detail. (A) Visualization of how data flows through the cell, where users can inspect how the cell state gets updated and how output activations are computed. (B) We show the input to the network and its prediction, and visualize how the training error is computed and then backpropagated through the network. A grey sliding window indicates which data points are needed to initialize the cell state. (C) The explanations now have more detailed steps for the forward direction of data flow. (D) Users can, in addition to interactively training the network, also change the speed in which the visualization for cell steps advances. (E) Just as in the network overview, users can modify training parameters.

an abstract, two-dimensional distribution of points to train their networks on. With *explorNN*, we follow this approach of employing rather simple data (V1, Complexity). However, as RNNs are focused on sequential data, we decided to use periodical mathematical functions. For our function data, the x-dimension is the periodicity of the function, and maps well to the temporal concept of RNN inputs (V2, Dynamics). The y-dimension is then the function value at a point in this timeline. The functions we provide as possible training inputs can be seen in Table 1, and vary in their periodicity, which means that the network has to learn this property from the training data. On top of this, users can add random noise to the functions, to make the training objective more complex (V1, Complexity).

In contrast to the aforementioned approaches, where only simplistic plots can be learned, we also provide more realistic next character prediction text training data. To allow for interactive training of these text-based data sets, we, again, employ rather simplistic text samples (V1, Complexity). These include a recurring character sequence (*ababab...*) and a simple text, the well-known *lorem ipsum*.

To visually communicate how the network operates on such sequential data, we use animation throughout the visualizations we propose. Accordingly, to demonstrate this data-dynamics, we animate the selected input function, such that it seems to flow while being input to the network (V2, Dynamics). Additionally, the prediction values also slowly build up with animations, to clarify that this prediction is building up sequentially, rather than consisting only of one prediction (V2, Dynamics).

On top of that, the sequential nature of RNNs is communicated in *explorNN* through textual descriptions and guiding the user directly to the input functions when first opening the application.

### 6.3 Training

To show the mechanism of backpropagation in our visualizations (O3), we communicate the direction of data flow throughout the different steps of the network training process. The connections in the network

follow this line, as they are animated with the flow of data. Here, dashed lines are moving forward during the prediction phase, which tightly integrates with the visualization of our input data and output predictions (V2, Dynamics).

To clarify that RNNs use backpropagation for weight updates, and that this backpropagation is based on the prediction error, we again exploit animation. To first illustrate the error calculation, we use vertical lines in the function plots, which slowly emerge between the prediction and the target value during error calculation (V2, Dynamics). This, again, visualizes the connection between data and processing inside the network, as it happens between the prediction and backpropagation phases of the training process (V2, Dynamics). After this animation, which visualizes how the error for an RNN is computed, is completed, we then reverse the animation of the connections between individual layers in the network overview, or layer components in the cell view, to illustrate that the data-flow direction is now reversed (V2, Dynamics).

Apart from backpropagation through layers, RNNs also backpropagate the error through time, which is visualized in the network overview (Fig. 1) through looping connections attached to each network layer. When inspecting individual cells (Fig. 3), the data flow of these reverse operations can be inspected in even more detail, as here, the animations for the connections between individual cell components are now reversed for backpropagation (V3, Multiscale). Within a cell, users can also see that this reverse data flow not only happens once to backpropagate to the previous layer, but multiple times, for all the input time steps (V2, Dynamics).

### 6.4 LSTM Cells

To both, show an overview of the training process and give detailed insight into the computation that is performed within one recurrent cell (O4), we employ an overview first, zoom and filter, details on demand visualization design, following Shneiderman’s mantra [28].

From the network overview visualization (Fig. 1), which shows an abstract visualization of the network layers, users can select one of the

layers, to investigate how an individual LSTM cell functions (Fig. 3) (V3, Multiscale).

Within this cell view, the LSTM visualization (Fig. 2) serves as a detailed explanation for an LSTM cell, as it shows how data flows through a cell (V3, Multiscale). The data-flow within these cells is visualized with step-by-step animations of the cell processes, where functions and connections that currently contain data relevant for processing subsequent steps are highlighted (V2, Dynamics). This way, we are able to communicate how the hidden state and output of these cells is computed, and visualize how the data flows from one compute unit to the next operation or gate. By showing the connections within the cell, it also becomes clear that there are two recurrent cycles, one from the output of the cell back to the input, and one within the cell, which is used to update the cell state based on its state in the previous iteration. While all the gates which transform the data are depicted with circular icons, the cell state, which, in contrast, represents the saved state of the cell, is depicted by a squared icon, namely a memory card.

Orientation in this multiscale approach is of great importance. We, therefore, employ a color-coding, which always indicates the current level of detail (V3, Multiscale). This highlight color, which signals the state of interactive components throughout the application, is blue for the network overview, whereas for the cell view, orange is used. Orange and blue are complementary colors, which makes them easily distinguishable, and they can additionally be differentiated by vision-impaired users [16]. To additionally support the user’s mental image of looking into one of the layers (V3, Multiscale), we employ a zooming transition onto one of the network layers, which gradually reveals the structure of a LSTM cell.

On top of that, each cell component can be selected to get a detailed explanation of its functionality, as shown in Fig. 5, which marks another level of detail in this visualization (V3, Multiscale).

## 6.5 Task

We also intend to communicate to users that the knowledge they obtained is transferable to more realistic problem contexts (O5). This is why, in addition to abstract function continuation, we also provide text-based data to train the network on (V1, Complexity). To do so, we employ a similar design language, as in the function learning scenario, such that users directly see that most ideas can be transferred across tasks (V1, Complexity). The main difference is the data-plot visualization, where predictions are displayed. Here, we instead show how individual characters are predicted by the network, and how these, based on this text continuation can be learned. This directly relates to the idea of predicting the next function value, in the function learning scenario. By incorporating this text learning scenario, users of *exploRNN* get to learn and inspect not only abstract problems, but can also learn based on more realistic scenarios (V1, Complexity).

## 7 INTERACTION DESIGN

Within *exploRNN*, the visualized network can be interactively trained by the user. In the following, we first describe general interaction concepts improving *exploRNN*’s usage, before elaborating about interactions in the different views of our environment in the upcoming subsections.

**Onboarding.** Interactive systems often have the problem that users

Table 1. Functions used to generate training data for our application. We provide a sinusoid function, a sawtooth function, an oscillating function, and a composite sinusoid function. All functions can also be overlaid with random noise to make the training task more complex.

Function	Period
 $f(x) = \sin(x)$	$2\pi$
 $f(x) = -1 + 2 \cdot \left(\frac{x \bmod \pi}{\pi}\right)$	$\pi$
 $f(x) = \text{sgn}\left(\sin\left(\frac{\pi}{2}x\right)\right)$	4
 $f(x) = \left(\sin\left(\frac{3}{2}x\right) + \sin\left(\frac{9}{2}x\right)\right) \cdot \frac{2}{3}$	$\frac{4}{3}\pi$

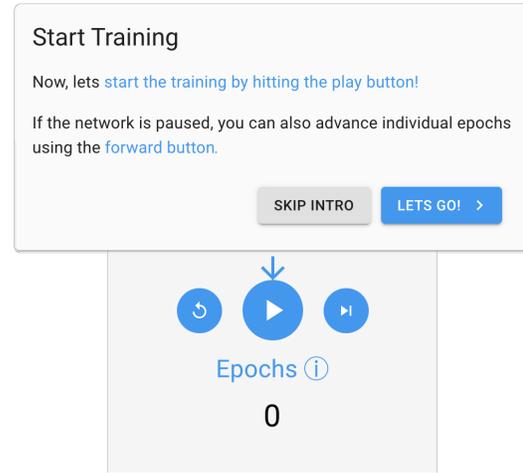


Fig. 4. Onboarding dialogs like this are shown when a user first launches *exploRNN*. They guide the user through the visualizations, so that no manual introduction is needed, and users can explore *exploRNN* on their own. Onboarding dialogs are tightly integrated into our visualizations, and use arrows to show users what is currently described.

need an introduction before being able to use them. While we designed *exploRNN* in a way that allows exploration without running the risk of making irreversible errors, we also propose to use an onboarding process for our educational environment. One step of this onboarding process is shown in Fig. 4, and helps users discover the capabilities of *exploRNN* (V1, Complexity).

**Textual Explanations.** In contrast to other learning environments, which show static textual explanations below the main visualization [17, 29], we instead provide textual explanations upon interaction (V3, Multiscale). This way, users can access more information for exactly those components they want to learn more about, while not having to read through a lot of text to find this information. Our interactively explorable dialog boxes integrate seamlessly with our proposed visualizations, and provide information about all important elements of the learning environment. An example for these dialogs can be seen in Fig. 5. Such dialogs exist for all headings, and are anchored through an information icon. Additionally, the components of an LSTM cell which are shown in the cell view, also each provide such explanations, as is alluded to in our onboarding process.

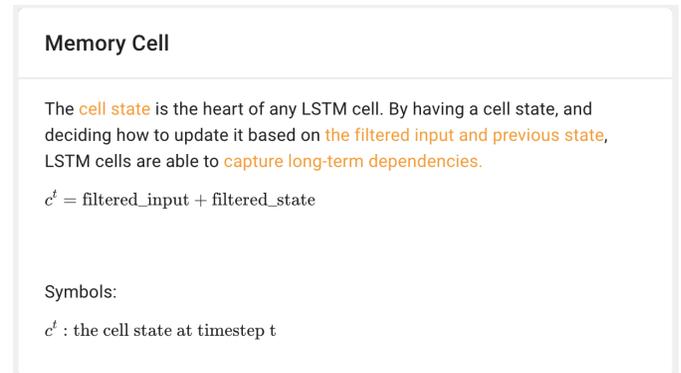


Fig. 5. Upon selecting components in our visualizations, we show detailed explanations. These explanation boxes are overlaid on the visualization and explain aspects of the visualization in detail. They are available for many items in our visualization, such as training steps, hyper-parameters, and operations in a cell.

## 7.1 Network Interaction

In the following, we are going to explain the interaction design for the network overview. For this, we will go over all the visualization components, whereas we use the letters A through F, as shown in Fig. 1, to indicate which component we currently discuss.

**A: Input.** we follow the concept of simplistic data and present visualizations of the functions and text samples that we proposed in Sect. 6.2. To experiment with the network, users can change the input data that is used for training the network here (V1, Complexity). This can help them understand what kind of data a RNN can process.

**B: Network.** Importantly, users can also investigate how the training progresses differently depending on the number of recurrent layers in the network. To help users explore this, they can add or remove layers from the network to be trained by simply interacting with the network visualization as shown in Fig. 6. Changing the network architecture automatically resets the training process, so that users can train this new network from scratch.

**C: Predictions.** Commencing the top row of visualizations to the right, is the output plot, where we visualize the prediction of the network for an evaluation sample alongside its ground truth. Based on this prediction, we then visualize the error by calculating the difference between the prediction and the ground truth value. This animated component, while not being interactive itself, marks an important feature of our environment, with which the results of interactions that have been made in other places can be inspected (V2, Dynamics).

**D: Process.** We split the training process into three visually distinct steps: prediction, validation, and training. The explanation pane in the lower left of the network overview (see Fig. 1) both, displays, which step is currently executed, and provides an explanation for what happens in each of these steps. Through this, the user can learn more about the training dynamics of the network.

**E: Controls.** In the network overview, the network is trained by means of epochs, which follows the "big-picture" approach of this visualization (V3, Multiscale). To experiment with this training process, users can interact with the control area in the bottom center of our environment. Apart from automatically advancing epochs, which can be done with the play button, users can also trigger the training of the network for a single epoch, by pressing the forward button, which gives users even more control over the training process (V2, Dynamics). The training process can additionally always be reset, using the rewind button.

**F: Training Parameters.** Experimenting with parameters for the training process is further supported by means to manipulate the learning rate and the batch size for the optimization, as these are critical as well as accessible parameters of the optimization process (V1, Complexity). Additionally, to make the training data more realistic, users can add noise to the input of the network, which more closely resembles real-world scenarios of imperfect measurements (V1, Complexity). These parameter changes can be made through sliders, which are positioned on the bottom right.

Following Shneiderman's mantra [28], the user can then, after getting an overview of the network, inspect individual LSTM cells in detail by selecting one of the layers in the network view to dive into the cell view (V3, Multiscale).

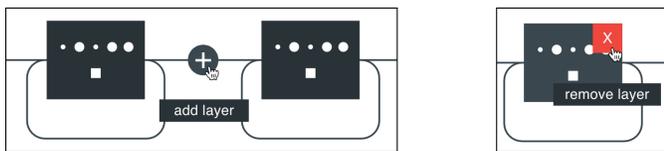


Fig. 6. left: Adding a layer between two existing layers. When the user hovers over the connection between two layers, a button to add another layer appears. right: Removing a layer from the network. Hovering over a layer in the network will show a button to remove that layer. Whenever the layers in the network are changed, training is reset to be directly started with for the new network architecture.

## 7.2 LSTM Cell Interaction

This subsection will cover the interaction design for the cell view. Again, we will go over all the visualization components, whereas we use the letters A through E, as shown in Fig. 3, to indicate which component we currently discuss

**A: Cell architecture.** To convey the functionality of one recurrent unit, we visualize how the cell input is fetched from the previous layer, and then gets processed by the compute functions within the recurrent cell (V2, Dynamics; V3, Multiscale). The components within this cell visualization, as well as the connections between them, are highlighted when activated, in order to provide feedback on the interactive training process (V2, Dynamics). More explanation, and equations for the computations within these cell components can be obtained by selecting one of the components within the cell visualization, providing more information only where needed (V1, Complexity; V3, Multiscale).

**B: Data Plot.** To the right of the cell visualization, we present the input data, network prediction, and ground truth all in one graph, disentangled from the cell visualization. As this cell typically receives data from previous cells, and outputs data to subsequent cells, we do not visualize the input to the left, and the output to the right of the cell, as we do in the network overview. By clearly separating the training data and predictions from the cell visualization, and embedding the cell between pictograms of surrounding cells, we are able to convey that the cell is not immediately receiving the input data, or directly making predictions (V2, Dynamics). Users can inspect this view during the interactively controlled training to see how the network processes input data to make predictions and calculate the training loss in relation to the processing steps within a cell (V2, Dynamics).

**C: Training Process.** The three steps of predicting, validating the prediction, and training the network are just as relevant in the cell view as they are in the network overview. However, here, they can also be directly jumped to, allowing the user to skip part of the data processing within a cell, and go directly to the processing step of interest (V2, Dynamics). This is only needed in the cell view, as one epoch is taking much longer here because we advance by compute steps rather than epochs (V3, Multiscale). For the forward pass, we additionally add explanations for the different processing steps of *receiving the layer input, calculating the gate activations, updating the cell state, and outputting the activation value* (V2, Dynamics). Users can use these explanations to visually map these recurring processing steps during the forward pass to the data flow visualized above.

**D: Controls.** In the cell view, processing is done by means of compute steps, showing a much more detailed processing pipeline than in the network overview (V3, Multiscale). Here, users can, again, use the control area in the bottom center of the environment to experiment with the training process. The more fine-grained advancement of the visualization is also adopted by how much the animation advances with the forward button, as it only executes the next compute step within a cell. In addition to what can be done in the network overview, the speed of the animations for data processing within this cell can be adjusted, so that users can explore the processing steps at their own pace (V2, Dynamics).

**E: Training Parameters.** Training parameters can be adjusted in the cell view just as in the network overview, to give the user more control over the training process.

To get back to the network overview, users can simply click anywhere outside the LSTM visualization in Fig. 3, A (V3, Multiscale).

## 8 TECHNICAL REALIZATION

While the visualization design described above, has been carefully crafted to meet the visualization design challenges outlined in Sect. 5.1, its technical realization needs to take into account the technical challenges identified in Sect. 5.2. In this section, we detail how we tackled these technical challenges.

### 8.1 Training Time

While a RNN for a complex application could not be trained live in the browser, we simplify the problem in multiple ways. First, by employing simplistic data sets, as described above, the model can converge after

relatively few epochs. This is important, as the user should not be required to train for a long time to be able to see improvements in prediction. Additionally, we limit the data points that are fed to the network per epoch, so that epochs are processed fast enough for our interactive visualization approach. We also limit the network size to at most seven layers, so that memory consumption and processing time are reduced. All these measures together enable a responsive, interactive learning environment, in which an RNN can be trained interactively. Thus, users can see the training progress and get visible prediction improvements after only a few epochs, while one such epoch takes seconds to compute in a current browser.

## 8.2 Training Steps

A key aspect of a controllable system is the decoupling of computation and visualization. To make training inspectable, we always run one training epoch, whereafter we fetch a prediction from the network. This prediction is then used to guide our visualization. Users of *explorNN* can additionally inspect the training process at their own pace, by progressing through training in a step-by-step manner. Through the decoupling of computation and visualization, we are able to show the training steps at an observable pace. This helps users understand how the model processes input data and predicts new data points.

## 8.3 Deployment

To be able to make *explorNN* publicly available for a large audience, we implemented it as an interactive browser application using HTML and JavaScript. We use TensorFlowJS, a JavaScript library for TensorFlow development [30] for training the RNN. For animated visualizations of the trained network, we use P5.js, a javascript library for processing [21]. This way, we are able to provide an interactive, web application that visualizes the training dynamics of RNNs through animation, which is accessible at: <https://papersubmissions42.github.io/explorNN/>. To help users learn about RNNs without the help of a teacher, we employ an onboarding process, as explained in Sect. 7.

# 9 USE CASES

In the following, we describe exemplar use cases for *explorNN*, in order to demonstrate the envisioned workflow, of how users can learn about RNNs with the help of our explorative visualization approach.

## 9.1 Learning about RNNs

Assume, that Bob is a student at a University. In a course on computer vision, he learned about MLPs and CNNs, and understands the fundamental concepts of deep learning well. Now, he is doing a student project on temporal data in the medical domain. He has recordings of breathing patterns, which he wants to predict the continuation of. His supervisor told him that RNNs are good at handling temporal data. Excited about his project, he uses *explorNN* to learn about the fundamental concepts of RNNs.

As Bob opens *explorNN*, he is greeted with the onboarding process, and is delighted to find out, that RNNs are indeed what he was searching for, in that they can be used to predict sequence continuation. After looking at the network architecture in the network overview visualization, he finds out that the overall structure of RNNs is not that different from CNNs, and decides to experiment with the training process.

After training the network for a while, and finding out that he can train a network that continues a periodic function, he wants to look into what makes RNN cells special, and why layers are depicted with a loop, and is guided by the onboarding process to enter the cell view. Now he can examine the training process in great detail, and finds out that RNNs are somehow capable of building up an internal state from previous inputs to predict the continuation of a sequence. Bob wonders how this is possible, and now turns his attention to the LSTM cell visualization. He is told that there is more information for all the components in *explorNN*, so he selects the LSTM cell heading, to get more information on how sequences are processed.

He learns that, by feeding back the activation of the previous time step, an RNN can include information about past data points into its

prediction process. The only thing that Bob still does not understand, is how further predictions can be made, where he does not have ground truth data anymore. He thus selects the network data visualization to the right of the LSTM cell, and now learns that past predictions can be used to build up the internal state further.

The next time Bob meets his supervisor, he tells him that he understands the basic concepts of RNNs, and will now look into implementing them for his problem.

## 9.2 Comprehending LSTM Cell Mechanics

Assume, Susan is a computer science student, who just attended a deep learning lecture, and understands the basic concepts behind RNNs. She heard that some RNN architectures suffer from the vanishing gradient problem, but also that LSTMs are designed to circumvent this problem. *explorNN* was briefly shown in the lecture she was attending, and the professor mentioned that it visualizes the widely used LSTM architecture. To find out how LSTMs are different from plain recurrent cells, Susan launches *explorNN*.

When looking at the cell view, which visualizes an individual LSTM cell, Susan directly notices that there are many more elements in this cell than in the vanilla RNN cell structure she knows about. She lets the network process some data, and notices, that the input to the cell consists of the input of the previous cell, as well as this cell's previous activation, which she is familiar with from vanilla RNN cells. However, she sees that LSTM cells contain three gates to filter this information.

Upon investigating the data-flow through these gates, Susan notices that two of them are used for a memory element, which LSTM cells possess. She investigates this element in more detail through our description panels of these individual components, and sees that it is updated with new data and the previous cell state. Interested in how this update is processed, she investigates the input and forget gates, and understands that through these, the network can learn to incorporate only relevant new information, and select information to be forgotten, which is not important anymore.

After looking at this internal cell state, and how it is updated, Susan now understands how LSTM cells can capture long-range dependencies, which vanilla RNN cells often cannot. She further looks at how the cell state is used to compute an output activation, and learns that the third gate, the output gate, is used to filter the internal memory for relevant information based on the current time step.

Susan used *explorNN* to investigate all the components of an LSTM cell, and took advantage of our visualization approach, where individual components can be investigated in great detail. She now knows how LSTM cells can outperform vanilla RNN cells, and that this comes at the cost of a few more processing steps. The next time she is confronted with the vanishing gradient problem for RNNs, she knows that using LSTM cells might help mitigate this problem.

# 10 USER STUDY

To evaluate the effectiveness of our approach, as well as our visual design, we conducted a qualitative user study with 6 participants (male: 5 part., female: 1 part.) aged between 23 and 32, via individual teleconferences<sup>1</sup>. All participants had some prior knowledge about CNNs, but were no experts in deep learning.

## 10.1 Setup

Our study consisted of a short demographic questionnaire, which was followed by self-evaluation questions about the knowledge concerning RNNs. These results were used as an indication of the prior knowledge of the participants. The participants then used *explorNN* for 15 minutes, whereby we gave them the following three goals: first, trying to understand how the network processes data to make predictions and how it can be trained based on that, second, trying to understand how individual network cells function, and third, comparing these insights from function data to text-based predictions. These goals were set, as the users of our study did not have an intrinsic motivation to learn

<sup>1</sup>Due to the closings caused by the Coronavirus spreading, we were not able to conduct a lab study with more participants.

these things, in contrast to typical users of our visualization approach. We, thus, think that these goals help well to reflect a typical scenario where a user wishes to learn about RNNs with the help of our system. Other than this guidance, the users did not get any introduction by us, but were entirely guided by our onboarding mechanism. After this interactive session of using *exploRNN*, users had to answer another self-evaluation on their knowledge about RNNs, questions about their learning preferences, and subsequently answered some questions about their experience with the proposed learning environment, which included answering the SUS [4] questionnaire. We concluded our study with a short semi-structured interview (10-15 minutes). The whole study took between 45 minutes and one hour per participant, and participants received a 10\$ Amazon gift card as compensation.

## 10.2 Results

In the following, we will present the results of our user study, grouped in the individual question types that participants had to answer.

**Self-Evaluation Questionnaire.** When asked about RNNs before using *exploRNN*, self estimates of knowledge were between one (2 part.) and two (3 part.), in addition to only one three (1 part.) (1: No knowledge, 5: Expert). Additionally, only three of our six participants were able to even name an application of RNNs. This indicates that most participants were not experienced with RNNs before using our environment. After using *exploRNN*, the self estimate about the knowledge of RNNs increased for all participants, except one, by one or two points, landing between two (2 part.) and three (3 part.), with one additional four (1 part.). Additionally, now all our participants could name an application for RNNs. Interestingly, all participants claimed that the learning effect was between four (4 part.) and five (1 part.) (1: learned nothing, 5: learned a lot).

**Learning Preferences.** When asked which kind of learning environment they preferred, all participants reported to prefer *exploRNN* to reading a blog post about the topic, and either valued *exploRNN* on par with a lecture about RNNs (3 part.), or slightly better (3 part.). All participants also thought that they would learn more from *exploRNN* than from a blog post, and all except one thought they would learn equally much (2 part.) or more (3 part.) than in a lecture. Most of the participants would, however, have liked to get even more information from *exploRNN*.

**System Usability.** Regarding the SUS questionnaire, the usability of our learning environment scored 84.58 points (SD: 2.71), which results in *excellent* usability [1]. This is supported by all except one participant assigning the system a five in terms of visual design (1: did not like it, 5: liked it a lot).

**Interviews.** The semi-structured interview contained questions roughly grouped into four categories, first, knowledge change from before to after using the system, second, questions to evaluate if the participants really learned something about RNNs, third, on which level of prior knowledge and interest *exploRNN* helps most, and fourth, questions about the visual design of *exploRNN*.

Regarding the first of these groups, all participants were sure, that they learned from the system (e.g.: *the hierarchic visualization [...] helped enormously*, and: *looking into what happens when and how helped me a lot more than the static image I saw in a lecture*), while some mentioned that, if given more time, they would have been able to learn even more than in just 15 minutes.

When we asked the participants specific questions about what they learned, it became clear that their knowledge indeed increased, as they now all could explain on an abstract level, e.g., what gates an LSTM consists of, and knew, e.g., that previous predictions were used to build up the internal cell state for future predictions, which they before did not know.

On the level on which *exploRNN* operates, all commented that is was highly useful for getting an overview. Some also thought that the mathematical details and the provided interactions allowed for more than that. However, all agreed, that for a more detailed system, which could be used to learn all the mathematical details or even implementation approaches, there would need to be more information (e.g.: *for being able to implement something like this, information is still missing*).

Many mentioned along this line, that if more information would be provided, this would come as a trade-off, where people that are not already experienced in this area, might, in turn, suffer from reduced simplicity. Regarding the type of learner who would benefit most of the environment, all participants agreed that people with some prior knowledge about deep learning, but who are new to sequence learning, would benefit most of using *exploRNN* (e.g.: *for experts and professionals, this could be too much of an overview*, and: *I think this would be good for motivating people who are new to the area and want to learn more about it*, or: *if someone has an idea of how other networks work, this application can be used to build upon that knowledge*).

When asked about the visual design of *exploRNN*, all participants expressed to have liked its simplicity and visual appeal (e.g.: *the learning process is nicely illustrated through the interactions that happen*), which supports the results of the SUS questionnaire. They also gave some comments for improvements, of which we already implemented some, such as replacing the blank space character for text prediction. Other comments, such as integrating visualizations for the loss function, or showing how a letter flows through the computations of a cell, were not implemented as we think this would harm the simplicity of the environment, and make understanding on a high level much more difficult (i.e. another, complicated parameter that would need to be explained and complex transformation into embedding space before flowing through the network), while not compensating for that with the benefits it could bring.

All in all, our study indicates that users can learn about RNNs with the help of our interactive environment. Furthermore, most participants even think that using *exploRNN* has a great learning effect, and most participants prefer it when compared to blog posts or lectures. While we got helpful feedback about possible improvements, the semi-structured interviews, as well as the usability evaluation suggest, that our system is well-designed, and that the visualization concept helps users to learn about RNNs.

## 11 CONCLUSIONS AND FUTURE WORK

In this paper, we propose a learning environment tailored towards educating about RNNs. To inform our visualization design, we first introduced educational objectives for learning about RNNs. Based on this, we identified design challenges for such an interactive learning environment, which we tackle in the proposed learning environment. Within this environment, we propose a new visualization approach for inspecting RNNs, where different levels of granularity are employed. The results of our qualitative user study indicate, that *exploRNN* indeed helps with learning about RNNs, and is even considered superior to other forms of learning by many participants. While there exist learning environments for other network types, to our knowledge, this is the first that covers the important area of sequence processing. For broadly accessible education, *exploRNN* can be used in any modern browser at <https://papersubmissions42.github.io/exploRNN/>.

Remaining open challenges for the visualization community with regards to educating about RNNs are, e.g., the inclusion of more RNN cell types into the learning environment. Additionally, it would be interesting to look into more types of tasks for RNNs. In our interviews, participants mentioned that they would like similar environments to learn about how to implement such networks. While current visualization approaches mainly provide an overview of these techniques, implementation is still done mostly with non-interactive, text-based tutorials. Interactive visualization tools, that teach users how to implement deep learning techniques could further ease entering the attractive field of deep learning research.

## REFERENCES

- [1] A. Bangor, P. Kortum, and J. Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123, 2009.
- [2] Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.

- [3] D. Britz, A. Goldie, M.-T. Luong, and Q. Le. Massive exploration of neural machine translation architectures. *arXiv preprint arXiv:1703.03906*, 2017.
- [4] J. Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.
- [5] D. Cashman, G. Patterson, A. Mosca, N. Watts, S. Robinson, and R. Chang. Rnnbow: Visualizing learning via backpropagation gradients in rnns. *IEEE computer graphics and applications*, 38(6):39–50, 2018.
- [6] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
- [7] S. Chung, S. Suh, C. Park, K. Kang, J. Choo, and B. C. Kwon. Revacnn: Real-time visual analytics for convolutional neural network. In *KDD 16 Workshop on Interactive Data Exploration and Analytics*, 2016.
- [8] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber. A novel connectionist system for unconstrained handwriting recognition. *IEEE transactions on pattern analysis and machine intelligence*, 31(5):855–868, 2008.
- [9] A. Graves, A.-r. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pp. 6645–6649. IEEE, 2013.
- [10] A. W. Harley. An interactive node-link visualization of convolutional neural networks. In *International Symposium on Visual Computing*, pp. 867–877. Springer, 2015.
- [11] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [12] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [13] S. Hochreiter and J. Schmidhuber. Lstm can solve hard long time lag problems. In *Advances in neural information processing systems*, pp. 473–479, 1997.
- [14] C. D. Hundhausen and J. L. Brown. What you see is what you code: A “live” algorithm development and visualization environment for novice learners. *Journal of Visual Languages & Computing*, 18(1):22–47, 2007.
- [15] C. D. Hundhausen, S. A. Douglas, and J. T. Stasko. A meta-study of algorithm visualization effectiveness. *Journal of Visual Languages & Computing*, 13(3):259–290, 2002.
- [16] B. Jenny and N. V. Kelso. Color design for the color vision impaired. *Cartographic perspectives*, (58):61–67, 2007.
- [17] M. Kahng, N. Thorat, D. H. P. Chau, F. B. Viégas, and M. Wattenberg. Gan lab: Understanding complex deep generative models using interactive visual experimentation. *IEEE transactions on visualization and computer graphics*, 25(1):1–11, 2018.
- [18] A. Karpathy. ConvnetJS mnist demo. <https://cs.stanford.edu/people/karpathy/convnetjs/demo/mnist.html>, 2020.
- [19] A. Karpathy, J. Johnson, and L. Fei-Fei. Visualizing and understanding recurrent networks. *arXiv preprint arXiv:1506.02078*, 2015.
- [20] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.
- [21] L. McCarthy. P5.js. <https://p5js.org/>, 2020.
- [22] Y. Ming, S. Cao, R. Zhang, Z. Li, Y. Chen, Y. Song, and H. Qu. Understanding hidden memories of recurrent neural networks. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 13–24. IEEE, 2017.
- [23] A. P. Norton and Y. Qi. Adversarial-playground: A visualization suite showing how adversarial examples fool deep learning. In *2017 IEEE Symposium on Visualization for Cyber Security (VizSec)*, pp. 1–4. IEEE, 2017.
- [24] R. Pascanu, T. Mikolov, and Y. Bengio. On the difficulty of training recurrent neural networks. In *International conference on machine learning*, pp. 1310–1318, 2013.
- [25] H. Sak, A. W. Senior, and F. Beaufays. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. 2014.
- [26] D. Schweitzer and W. Brown. Interactive visualization for the active learning classroom. In *Proceedings of the 38th SIGCSE technical symposium on Computer science education*, pp. 208–212, 2007.
- [27] Q. Shen, Y. Wu, Y. Jiang, W. Zeng, A. K. LAU, A. Vianova, and H. Qu. Visual interpretation of recurrent neural network on multi-dimensional time-series forecast. *IEEE transactions on visualization and computer graphics*, 2020.
- [28] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE symposium on visual languages*, pp. 336–343. IEEE, 1996.
- [29] D. Smilkov, S. Carter, D. Sculley, F. B. Viégas, and M. Wattenberg. Direct-manipulation visualization of deep networks. *arXiv preprint arXiv:1708.03788*, 2017.
- [30] D. Smilkov, N. Thorat, Y. Assogba, A. Yuan, N. Kreeger, P. Yu, K. Zhang, S. Cai, E. Nielsen, D. Soergel, et al. Tensorflow.js: Machine learning for the web and beyond. *arXiv preprint arXiv:1901.05350*, 2019.
- [31] H. Strobelt, S. Gehrmann, M. Behrisch, A. Perer, H. Pfister, and A. M. Rush. Seq2seq-vis: A visual debugging tool for sequence-to-sequence models. *IEEE transactions on visualization and computer graphics*, 25(1):353–363, 2018.
- [32] H. Strobelt, S. Gehrmann, H. Pfister, and A. M. Rush. Lstmvis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. *IEEE transactions on visualization and computer graphics*, 24(1):667–676, 2017.
- [33] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.
- [34] G. Weiss, Y. Goldberg, and E. Yahav. On the practical computational power of finite precision rnns for language recognition. *arXiv preprint arXiv:1805.04908*, 2018.
- [35] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al. Google’s neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*, 2016.
- [36] A. Ynnerman, J. Löwgren, and L. Tibell. Exploratron: A new science communication paradigm. *IEEE computer graphics and applications*, 38(3):13–20, 2018.