Long Short-Term Memory

Sepp Hochreiter
Fakultät für Informatik, Technische Universität München, 80290 München, Germany

Jürgen Schmidhuber
IDSIA, Corso Elvezia 36, 6900 Lugano, Switzerland

Learning to store information over extended time intervals by recurrent backpropagation takes a very long time, mostly because of insufficient, decaying error backflow. We briefly review Hochreiter's (1991) analysis of this problem, then address it by introducing a novel, efficient, gradient-based method called long short-term memory (LSTM). Truncating the gradient where this does not do harm, LSTM can learn to bridge minimal time lags in excess of 1000 discrete-time steps by enforcing constant error flow through constant error carousels within special units. Multiplicative gate units learn to open and close access to the constant error flow. LSTM is local in space and time; its computational complexity per time step and weight is $O(1)$. Our experiments with artificial data involve local, distributed, real-valued, and noisy pattern representations. In comparisons with real-time recurrent learning, back propagation through time, recurrent cascade correlation, Elman nets, and neural sequence chunking, LSTM leads to many more successful runs, and learns much faster. LSTM also solves complex, artificial long-time-lag tasks that have never been solved by previous recurrent network algorithms.

1 Introduction

In principle, recurrent networks can use their feedback connections to store representations of recent input events in the form of activations (short-term memory, as opposed to long-term memory embodied by slowly changing weights). This is potentially significant for many applications, including speech processing, non-Markovian control, and music composition (Mozer, 1992). The most widely used algorithms for learning what to put in short-term memory, however, take too much time or do not work well at all, especially when minimal time lags between inputs and corresponding teacher signals are long. Although theoretically fascinating, existing methods do not provide clear practical advantages over, say, backpropagation in feed-forward nets with limited time windows. This article reviews an analysis of the problem and suggests a remedy.

The problem. With conventional backpropagation through time (BPTT; Williams & Zipser, 1992; Werbos, 1988) or real-time recurrent learning (RTRL; Robinson & Fallside, 1987), error signals flowing backward in time tend to (1) blow up or (2) vanish; the temporal evolution of the backpropagated error exponentially depends on the size of the weights (Hochreiter, 1991). Case 1 may lead to oscillating weights; in case 2, learning to bridge long time lags takes a prohibitive amount of time or does not work at all (see section 3).

This article presents long short-term memory (LSTM), a novel recurrent network architecture in conjunction with an appropriate gradient-based learning algorithm. LSTM is designed to overcome these error backflow problems. It can learn to bridge time intervals in excess of 1000 steps even in case of noisy, incompressible input sequences, without loss of short-time-lag capabilities. This is achieved by an efficient, gradient-based algorithm for an architecture enforcing constant (thus, neither exploding nor vanishing) error flow through internal states of special units (provided the gradient computation is truncated at certain architecture-specific points; this does not affect long-term error flow, though).

Section 2 briefly reviews previous work. Section 3 begins with an outline of the detailed analysis of vanishing errors due to Hochreiter (1991). It then introduces a naive approach to constant error backpropagation for didactic purposes and highlights its problems concerning information storage and retrieval. These problems lead to the LSTM architecture described in section 4. Section 5 presents numerous experiments and comparisons with competing methods. LSTM outperforms them and also learns to solve complex, artificial tasks no other recurrent net algorithm has solved. Section 6 discusses LSTM's limitations and advantages. The appendix contains a detailed description of the algorithm (A.1) and explicit error flow formulas (A.2).

2 Previous Work

This section focuses on recurrent nets with time-varying inputs (as opposed to nets with stationary inputs and fixed-point-based gradient calculations; e.g., Almeida, 1987; Pineda, 1987).

2.1 Gradient-Descent Variants. The approaches of Elman (1988), Fahlman (1991), Williams (1989), Schmidhuber (1992a), Pearlmutter (1989), and many of the related algorithms in Pearlmutter's comprehensive overview (1995) suffer from the same problems as BPTT and RTRL (see sections 1 and 3).

2.2 Time Delays. Other methods that seem practical for short time lags only are time-delay neural networks (Lang, Waibel, & Hinton, 1990) and Plate's method (Plate, 1993), which updates unit activations based on a

weighted sum of old activations (see also de Vries & Principe, 1991). Lin et al. (1996) propose variants of time-delay networks called NARX networks.

2.3 Time Constants. To deal with long time lags, Mozer (1992) uses time constants influencing changes of unit activations (deVries and Principe's 1991 approach may in fact be viewed as a mixture of time-delay neural networks and time constants). For long time lags, however, the time constants need external fine tuning (Mozer, 1992). Sun, Chen, and Lee's alternative approach (1993) updates the activation of a recurrent unit by adding the old activation and the (scaled) current net input. The net input, however, tends to perturb the stored information, which makes long-term storage impractical.

2.4 Ring's Approach. Ring (1993) also proposed a method for bridging long time lags. Whenever a unit in his network receives conflicting error signals, he adds a higher-order unit influencing appropriate connections. Although his approach can sometimes be extremely fast, to bridge a time lag involving 100 steps may require the addition of 100 units. Also, Ring's net does not generalize to unseen lag durations.

2.5 Bengio et al.'s Approach. Bengio, Simard, and Frasconi (1994) investigate methods such as simulated annealing, multigrid random search, time-weighted pseudo-Newton optimization, and discrete error propagation. Their "latch" and "two-sequence" problems are very similar to problem 3a in this article with minimal time lag 100 (see Experiment 3). Bengio and Frasconi (1994) also propose an expectation-maximazation approach for propagating targets. With *n* so-called state networks, at a given time, their system can be in one of only *n* different states. (See also the beginning of section 5.) But to solve continuous problems such as the adding problem (section 5.4), their system would require an unacceptable number of states (i.e., state networks).

2.6 Kalman Filters. Puskorius and Feldkamp (1994) use Kalman filter techniques to improve recurrent net performance. Since they use "a derivative discount factor imposed to decay exponentially the effects of past dynamic derivatives," there is no reason to believe that their Kalman filter-trained recurrent networks will be useful for very long minimal time lags.

2.7 Second Order Nets. We will see that LSTM uses multiplicative units (MUs) to protect error flow from unwanted perturbations. It is not the first recurrent net method using MUs, though. For instance, Watrous and Kuhn (1992) use MUs in second-order nets. There are some differences from LSTM: (1) Watrous and Kuhn's architecture does not enforce constant error flow and is not designed to solve long-time-lag problems; (2) it has fully connected second-order sigma-pi units, while the LSTM architecture's MUs

are used only to gate access to constant error flow; and (3) Watrous and Kuhn's algorithm costs $O(W^2)$ operations per time step, ours only $O(W)$, where W is the number of weights. See also Miller and Giles (1993) for additional work on MUs.

2.8 Simple Weight Guessing. To avoid long-time-lag problems of gradient-based approaches, we may simply randomly initialize all network weights until the resulting net happens to classify all training sequences correctly. In fact, recently we discovered (Schmidhuber & Hochreiter, 1996; Hochreiter & Schmidhuber, 1996, 1997) that simple weight guessing solves many of the problems in Bengio et al. (1994), Bengio and Frasconi (1994), Miller and Giles (1993), and Lin et al. (1996) faster than the algorithms these authors proposed. This does not mean that weight guessing is a good algorithm. It just means that the problems are very simple. More realistic tasks require either many free parameters (e.g., input weights) or high weight precision (e.g., for continuous-valued parameters), such that guessing becomes completely infeasible.

2.9 Adaptive Sequence Chunkers. Schmidhuber's hierarchical chunker systems (1992b, 1993) do have a capability to bridge arbitrary time lags, but only if there is local predictability across the subsequences causing the time lags (see also Mozer, 1992). For instance, in his postdoctoral thesis, Schmidhuber (1993) uses hierarchical recurrent nets to solve rapidly certain grammar learning tasks involving minimal time lags in excess of 1000 steps. The performance of chunker systems, however, deteriorates as the noise level increases and the input sequences become less compressible. LSTM does not suffer from this problem.

3 Constant Error Backpropagation

3.1 Exponentially Decaying Error

3.1.1 Conventional BPTT (e.g., Williams & Zipser, 1992). Output unit k's target at time t is denoted by $d_k(t)$. Using mean squared error, k's error signal is

$$\vartheta_k(t) = f_k'(net_k(t))(d_k(t) - y^k(t)),$$

where

$$y^j(t) = f_i(net_i(t))$$

is the activation of a noninput unit i with differentiable activation function f_i,

$$net_i(t) = \sum_j w_{ij} y^j(t-1)$$

is unit i's current net input, and w_{ij} is the weight on the connection from unit j to i. Some nonoutput unit j's backpropagated error signal is

$$\vartheta_j(t) = f'_j(net_j(t)) \sum_i w_{ij} \vartheta_i(t+1).$$

The corresponding contribution to w_{jl}'s total weight update is $\alpha \vartheta_j(t) y^l(t-1)$, where α is the learning rate and l stands for an arbitrary unit connected to unit j.

3.1.2 Outline of Hochreiter's Analysis (1991, pp. 19–21). Suppose we have a fully connected net whose noninput unit indices range from 1 to n. Let us focus on local error flow from unit u to unit v (later we will see that the analysis immediately extends to global error flow). The error occurring at an arbitrary unit u at time step t is propagated back into time for q time steps, to an arbitrary unit v. This will scale the error by the following factor:

$$\frac{\partial \vartheta_v(t-q)}{\partial \vartheta_u(t)} = \begin{cases} f'_v(net_v(t-1)) w_{uv} & q = 1 \\ f'_v(net_v(t-q)) \sum_{l=1}^{n} \frac{\partial \vartheta_l(t-q+1)}{\partial \vartheta_u(t)} w_{lv} & q > 1 \end{cases}. \tag{3.1}$$

With $l_q = v$ and $l_0 = u$, we obtain:

$$\frac{\partial \vartheta_v(t-q)}{\partial \vartheta_u(t)} = \sum_{l_1=1}^{n} \cdots \sum_{l_{q-1}=1}^{n} \prod_{m=1}^{q} f'_{l_m}(net_{l_m}(t-m)) w_{l_m l_{m-1}} \tag{3.2}$$

(proof by induction). The sum of the n^{q-1} terms $\prod_{m=1}^{q} f'_{l_m}(net_{l_m}(t-m)) w_{l_m l_{m-1}}$ determines the total error backflow (note that since the summation terms may have different signs, increasing the number of units n does not necessarily increase error flow).

3.1.3 Intuitive Explanation of Equation 3.2. If

$$|f'_{l_m}(net_{l_m}(t-m)) w_{l_m l_{m-1}}| > 1.0$$

for all m (as can happen, e.g., with linear f_{l_m}), then the largest product increases exponentially with q. That is, the error blows up, and conflicting error signals arriving at unit v can lead to oscillating weights and unstable learning (for error blowups or bifurcations, see also Pineda, 1988; Baldi & Pineda, 1991; Doya, 1992). On the other hand, if

$$|f'_{l_m}(net_{l_m}(t-m)) w_{l_m l_{m-1}}| < 1.0$$

for all m, then the largest product decreases exponentially with q. That is, the error vanishes, and nothing can be learned in acceptable time.

If f_{l_m} is the logistic sigmoid function, then the maximal value of f'_{l_m} is 0.25. If $y^{l_{m-1}}$ is constant and not equal to zero, then $|f'_{l_m}(net_{l_m})w_{l_m l_{m-1}}|$ takes on maximal values where

$$w_{l_m l_{m-1}} = \frac{1}{y^{l_{m-1}}} \coth\left(\frac{1}{2} net_{l_m}\right),$$

goes to zero for $|w_{l_m l_{m-1}}| \to \infty$, and is less than 1.0 for $|w_{l_m l_{m-1}}| < 4.0$ (e.g., if the absolute maximal weight value w_{max} is smaller than 4.0). Hence with conventional logistic sigmoid activation functions, the error flow tends to vanish as long as the weights have absolute values below 4.0, especially in the beginning of the training phase. In general, the use of larger initial weights will not help, though, as seen above, for $|w_{l_m l_{m-1}}| \to \infty$ the relevant derivative goes to zero "faster" than the absolute weight can grow (also, some weights will have to change their signs by crossing zero). Likewise, increasing the learning rate does not help either; it will not change the ratio of long-range error flow and short-range error flow. BPTT is too sensitive to recent distractions. (A very similar, more recent analysis was presented by Bengio et al., 1994.)

3.1.4 Global Error Flow.   The local error flow analysis above immediately shows that global error flow vanishes too. To see this, compute

$$\sum_{u:\ u\ \text{output unit}} \frac{\partial \vartheta_v(t-q)}{\partial \vartheta_u(t)}.$$

3.1.5 Weak Upper Bound for Scaling Factor.   The following, slightly extended vanishing error analysis also takes n, the number of units, into account. For $q > 1$, equation 3.2 can be rewritten as

$$(W_{u^T})^T F'(t-1) \prod_{m=2}^{q-1} (WF'(t-m))\ W_v\ f'_v(net_v(t-q)),$$

where the weight matrix W is defined by $[W]_{ij} := w_{ij}$, v's outgoing weight vector W_v is defined by $[W_v]_i := [W]_{iv} = w_{iv}$, u's incoming weight vector W_{u^T} is defined by $[W_{u^T}]_i := [W]_{ui} = w_{ui}$, and for $m = 1, \ldots, q$, $F'(t-m)$ is the diagonal matrix of first-order derivatives defined as $[F'(t-m)]_{ij} := 0$ if $i \neq j$, and $[F'(t-m)]_{ij} := f'_i(net_i(t-m))$ otherwise. Here T is the transposition operator, $[A]_{ij}$ is the element in the ith column and jth row of matrix A, and $[x]_i$ is the ith component of vector x.

Using a matrix norm $\|\cdot\|_A$ compatible with vector norm $\|\cdot\|_x$, we define

$$f'_{max} := \max_{m=1,\ldots,q} \{\|F'(t-m)\|_A\}.$$

For $\max_{i=1,\ldots,n}\{|x_i|\} \leq \|x\|_x$ we get $|x^T y| \leq n\,\|x\|_x\,\|y\|_x$. Since

$$|f'_v(net_v(t-q))| \leq \|F'(t-q)\|_A \leq f'_{max},$$

we obtain the following inequality:

$$\left| \frac{\partial \vartheta_v(t-q)}{\partial \vartheta_u(t)} \right| \ \leq \ n \, (f'_{\max})^q \, \|W_v\|_x \, \|W_{u^T}\|_x \, \|W\|_A^{q-2} \ \leq \ n \left(f'_{\max} \, \|W\|_A \right)^q .$$

This inequality results from

$$\|W_v\|_x \ = \ \|W e_v\|_x \ \leq \ \|W\|_A \, \|e_v\|_x \ \leq \ \|W\|_A$$

and

$$\|W_{u^T}\|_x \ = \ \|W^T e_u\|_x \ \leq \ \|W\|_A \, \|e_u\|_x \ \leq \ \|W\|_A,$$

where e_k is the unit vector whose components are 0 except for the kth component, which is 1. Note that this is a weak, extreme case upper bound; it will be reached only if all $\|F'(t-m)\|_A$ take on maximal values, and if the contributions of all paths across which error flows back from unit u to unit v have the same sign. Large $\|W\|_A$, however, typically result in small values of $\|F'(t-m)\|_A$, as confirmed by experiments (see, e.g., Hochreiter, 1991).

For example, with norms

$$\|W\|_A \ := \ \max_r \sum_s |w_{rs}|$$

and

$$\|x\|_x \ := \ \max_r |x_r|,$$

we have $f'_{\max} = 0.25$ for the logistic sigmoid. We observe that if

$$|w_{ij}| \leq w_{\max} < \frac{4.0}{n} \ \forall i, j,$$

then $\|W\|_A \ \leq \ n w_{\max} \ < \ 4.0$ will result in exponential decay. By setting $\tau := \left(\frac{n w_{\max}}{4.0} \right) < 1.0$, we obtain

$$\left| \frac{\partial \vartheta_v(t-q)}{\partial \vartheta_u(t)} \right| \leq n \, (\tau)^q .$$

We refer to Hochreiter (1991) for additional results.

3.2 Constant Error Flow: Naive Approach.

3.2.1 A Single Unit. To avoid vanishing error signals, how can we achieve constant error flow through a single unit j with a single connection to itself? According to the rules above, at time t, j's local error backflow is $\vartheta_j(t) = f'_j(net_j(t))\vartheta_j(t+1)w_{jj}$. To enforce constant error flow through j, we require

$$f'_j(net_j(t)) w_{jj} = 1.0.$$

Note the similarity to Mozer's fixed time constant system (1992)—a time constant of 1.0 is appropriate for potentially infinite time lags.[1]

3.2.2 The Constant Error Carousel.   Integrating the differential equation above, we obtain

$$f_j(net_j(t)) = \frac{net_j(t)}{w_{jj}}$$

for arbitrary $net_j(t)$. This means f_j has to be linear, and unit j's activation has to remain constant:

$$y_j(t+1) = f_j(net_j(t+1)) = f_j(w_{jj}y^j(t)) = y^j(t).$$

In the experiments, this will be ensured by using the identity function f_j : $f_j(x) = x, \forall x$, and by setting $w_{jj} = 1.0$. We refer to this as the constant error carousel (CEC). CEC will be LSTM's central feature (see section 4).

Of course, unit j will not only be connected to itself but also to other units. This invokes two obvious, related problems (also inherent in all other gradient-based approaches):

1. Input weight conflict: For simplicity, let us focus on a single additional input weight w_{ji}. Assume that the total error can be reduced by switching on unit j in response to a certain input and keeping it active for a long time (until it helps to compute a desired output). Provided i is nonzero, since the same incoming weight has to be used for both storing certain inputs and ignoring others, w_{ji} will often receive conflicting weight update signals during this time (recall that j is linear). These signals will attempt to make w_{ji} participate in (1) storing the input (by switching on j) and (2) protecting the input (by preventing j from being switched off by irrelevant later inputs). This conflict makes learning difficult and calls for a more context-sensitive mechanism for controlling write operations through input weights.

2. Output weight conflict: Assume j is switched on and currently stores some previous input. For simplicity, let us focus on a single additional outgoing weight w_{kj}. The same w_{kj} has to be used for both retrieving j's content at certain times and preventing j from disturbing k at other times. As long as unit j is nonzero, w_{kj} will attract conflicting weight update signals generated during sequence processing. These signals will attempt to make w_{kj} participate in accessing the information stored in j and—at different times—protecting unit k from being perturbed by j. For instance, with many tasks there are certain short-time-lag errors that can be reduced in early training stages. However,

---

[1] We do not use the expression "time constant" in the differential sense, as Pearlmutter (1995) does.

at later training stages, j may suddenly start to cause avoidable errors in situations that already seemed under control by attempting to participate in reducing more difficult long-time-lag errors. Again, this conflict makes learning difficult and calls for a more context-sensitive mechanism for controlling read operations through output weights.

Of course, input and output weight conflicts are not specific for long time lags; they occur for short time lags as well. Their effects, however, become particularly pronounced in the long-time-lag case. As the time lag increases, stored information must be protected against perturbation for longer and longer periods, and, especially in advanced stages of learning, more and more already correct outputs also require protection against perturbation.

Due to the problems set out, the naive approach does not work well except in the case of certain simple problems involving local input-output representations and nonrepeating input patterns (see Hochreiter, 1991; Silva, Amarel, Langlois, & Almeida, 1996). The next section shows how to do it right.

4 The Concept of Long Short-Term Memory

4.1 Memory Cells and Gate Units. To construct an architecture that allows for constant error flow through special, self-connected units without the disadvantages of the naive approach, we extend the CEC embodied by the self-connected, linear unit j from section 3.2 by introducing additional features. A multiplicative input gate unit is introduced to protect the memory contents stored in j from perturbation by irrelevant inputs, and a multiplicative output gate unit is introduced to protect other units from perturbation by currently irrelevant memory contents stored in j.

The resulting, more complex unit is called a memory cell (see Figure 1). The jth memory cell is denoted c_j. Each memory cell is built around a central linear unit with a fixed self-connection (the CEC). In addition to net_{c_j}, c_j gets input from a multiplicative unit out_j (the output gate), and from another multiplicative unit in_j (the input gate). in_j's activation at time t is denoted by $y^{in_j}(t)$, out_j's by $y^{out_j}(t)$. We have

$$y^{out_j}(t) = f_{out_j}(net_{out_j}(t)); \, y^{in_j}(t) = f_{in_j}(net_{in_j}(t));$$

where

$$net_{out_j}(t) = \sum_u w_{out_j u} y^u(t-1),$$
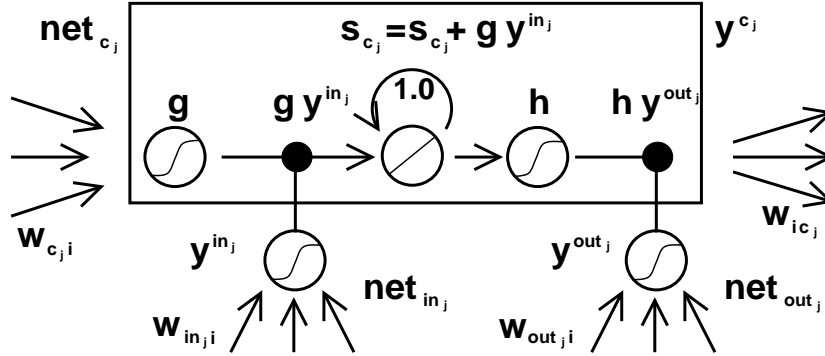
and

$$net_{in_j}(t) = \sum_u w_{in_j u} y^u(t-1).$$

Figure 1: Architecture of memory cell c_j (the box) and its gate units in_j, out_j. The self-recurrent connection (with weight 1.0) indicates feedback with a delay of one time step. It builds the basis of the CEC. The gate units open and close access to CEC. See text and appendix A.1 for details.

We also have

$$net_{c_j}(t) = \sum_u w_{c_j u} y^u(t-1).$$

The summation indices u may stand for input units, gate units, memory cells, or even conventional hidden units if there are any (see section 4.3). All these different types of units may convey useful information about the current state of the net. For instance, an input gate (output gate) may use inputs from other memory cells to decide whether to store (access) certain information in its memory cell. There even may be recurrent self-connections like $w_{c_j c_j}$. It is up to the user to define the network topology. See Figure 2 for an example.

At time t, c_j's output $y^{c_j}(t)$ is computed as

$$y^{c_j}(t) = y^{out_j}(t) h(s_{c_j}(t)),$$

where the internal state $s_{c_j}(t)$ is

$$s_{c_j}(0) = 0, s_{c_j}(t) = s_{c_j}(t-1) + y^{in_j}(t) g\left(net_{c_j}(t)\right) \text{ for } t > 0.$$

The differentiable function g squashes net_{c_j}; the differentiable function h scales memory cell outputs computed from the internal state s_{c_j}.

4.2 Why Gate Units? To avoid input weight conflicts, in_j controls the error flow to memory cell c_j's input connections $w_{c_j i}$. To circumvent c_j's output weight conflicts, out_j controls the error flow from unit j's output
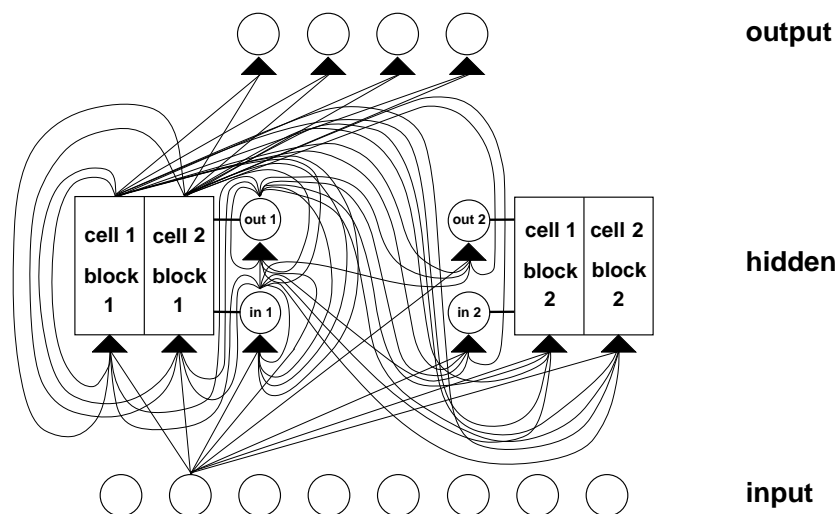
Figure 2: Example of a net with eight input units, four output units, and two memory cell blocks of size 2. *in*1 marks the input gate, *out*1 marks the output gate, and *cell*1/*block*1 marks the first memory cell of block 1. *cell*1/*block*1's architecture is identical to the one in Figure 1, with gate units *in*1 and *out*1 (note that by rotating Figure 1 by 90 degrees anticlockwise, it will match with the corresponding parts of Figure 2). The example assumes dense connectivity: each gate unit and each memory cell sees all non-output units. For simplicity, however, outgoing weights of only one type of unit are shown for each layer. With the efficient, truncated update rule, error flows only through connections to output unit, and through fixed self-connections within cell blocks (not shown here; see Figure 1). Error flow is truncated once it "wants" to leave memory cells or gate units. Therefore, no connection shown above serves to propagate error back to the unit from which the connection originates (except for connections to output units), although the connections themselves are modifiable. That is why the truncated LSTM algorithm is so efficient, despite its ability to bridge very long time lags. See the text and the appendix for details. Figure 2 shows the architecture used for experiment 6a; only the bias of the noninput units is omitted.

connections. In other words, the net can use in_j to decide when to keep or override information in memory cell c_j and out_j to decide when to access memory cell c_j and when to prevent other units from being perturbed by c_j (see Figure 1).

Error signals trapped within a memory cell's CEC cannot change, but different error signals flowing into the cell (at different times) via its output gate may get superimposed. The output gate will have to learn which

errors to trap in its CEC by appropriately scaling them. The input gate will have to learn when to release errors, again by appropriately scaling them. Essentially the multiplicative gate units open and close access to constant error flow through CEC.

Distributed output representations typically do require output gates. Both gate types are not always necessary, though; one may be sufficient. For instance, in experiments 2a and 2b in section 5, it will be possible to use input gates only. In fact, output gates are not required in case of local output encoding; preventing memory cells from perturbing already learned outputs can be done by simply setting the corresponding weights to zero. Even in this case, however, output gates can be beneficial: they prevent the net's attempts at storing long-time-lag memories (which are usually hard to learn) from perturbing activations representing easily learnable short-time-lag memories. (This will prove quite useful in experiment 1, for instance.)

4.3 Network Topology. We use networks with one input layer, one hidden layer, and one output layer. The (fully) self-connected hidden layer contains memory cells and corresponding gate units (for convenience, we refer to both memory cells and gate units as being located in the hidden layer). The hidden layer may also contain conventional hidden units providing inputs to gate units and memory cells. All units (except for gate units) in all layers have directed connections (serve as inputs) to all units in the layer above (or to all higher layers; see experiments 2a and 2b).

4.4 Memory Cell Blocks. *S* memory cells sharing the same input gate and the same output gate form a structure called a memory cell block of size *S*. Memory cell blocks facilitate information storage. As with conventional neural nets, it is not so easy to code a distributed input within a single cell. Since each memory cell block has as many gate units as a single memory cell (namely, two), the block architecture can be even slightly more efficient. A memory cell block of size 1 is just a simple memory cell. In the experiments in section 5, we will use memory cell blocks of various sizes.

4.5 Learning. We use a variant of RTRL (e.g., Robinson & Fallside, 1987) that takes into account the altered, multiplicative dynamics caused by input and output gates. To ensure nondecaying error backpropagation through internal states of memory cells, as with truncated BPTT (e.g., Williams & Peng, 1990), errors arriving at memory cell net inputs (for cell c_j, this includes net_{c_j}, net_{in_j}, net_{out_j}) do not get propagated back further in time (although they do serve to change the incoming weights). Only within memory cells, are errors propagated back through previous internal states s_{c_j}.[2] To visualize

---

[2] For intracellular backpropagation in a quite different context, see also Doya and Yoshizawa (1989).

this, once an error signal arrives at a memory cell output, it gets scaled by output gate activation and h'. Then it is within the memory cell's CEC, where it can flow back indefinitely without ever being scaled. When it leaves the memory cell through the input gate and g, it is scaled once more by input gate activation and g'. It then serves to change the incoming weights before it is truncated (see the appendix for formulas).

4.6 Computational Complexity. As with Mozer's focused recurrent back-propagation algorithm (Mozer, 1989), only the derivatives $\partial s_{c_j}/\partial w_{il}$ need to be stored and updated. Hence the LSTM algorithm is very efficient, with an excellent update complexity of $O(W)$, where W the number of weights (see details in the appendix). Hence, LSTM and BPTT for fully recurrent nets have the same update complexity per time step (while RTRL's is much worse). Unlike full BPTT, however, LSTM is local in space and time:[3] there is no need to store activation values observed during sequence processing in a stack with potentially unlimited size.

4.7 Abuse Problem and Solutions. In the beginning of the learning phase, error reduction may be possible without storing information over time. The network will thus tend to abuse memory cells, for example, as bias cells (it might make their activations constant and use the outgoing connections as adaptive thresholds for other units). The potential difficulty is that it may take a long time to release abused memory cells and make them available for further learning. A similar "abuse problem" appears if two memory cells store the same (redundant) information. There are at least two solutions to the abuse problem: (1) sequential network construction (e.g., Fahlman, 1991): a memory cell and the corresponding gate units are added to the network whenever the error stops decreasing (see experiment 2 in section 5), and (2) output gate bias: each output gate gets a negative initial bias, to push initial memory cell activations toward zero. Memory cells with more negative bias automatically get "allocated" later (see experiments 1, 3, 4, 5, and 6 in section 5).

4.8 Internal State Drift and Remedies. If memory cell c_j's inputs are mostly positive or mostly negative, then its internal state s_j will tend to drift away over time. This is potentially dangerous, for the $h'(s_j)$ will then adopt very small values, and the gradient will vanish. One way to circumvent this problem is to choose an appropriate function h. But $h(x) = x$, for instance, has the disadvantage of unrestricted memory cell output range. Our simple

---

[3] Following Schmidhuber (1989), we say that a recurrent net algorithm is *local in space* if the update complexity per time step and weight does not depend on network size. We say that a method is *local in time* if its storage requirements do not depend on input sequence length. For instance, RTRL is local in time but not in space. BPTT is local in space but not in time.

but effective way of solving drift problems at the beginning of learning is initially to bias the input gate in_j toward zero. Although there is a trade-off between the magnitudes of $h'(s_j)$ on the one hand and of y^{in_j} and f'_{in_j} on the other, the potential negative effect of input gate bias is negligible compared to the one of the drifting effect. With logistic sigmoid activation functions, there appears to be no need for fine-tuning the initial bias, as confirmed by experiments 4 and 5 in section 5.4.

5 Experiments

Which tasks are appropriate to demonstrate the quality of a novel long-time-lag algorithm? First, minimal time lags between relevant input signals and corresponding teacher signals must be long for all training sequences. In fact, many previous recurrent net algorithms sometimes manage to generalize from very short training sequences to very long test sequences (see, e.g., Pollack, 1991). But a real long-time-lag problem does not have any short-time-lag exemplars in the training set. For instance, Elman's training procedure, BPTT, offline RTRL, online RTRL, and others fail miserably on real long-time-lag problems. (See, e.g., Hochreiter, 1991; Mozer, 1992.) A second important requirement is that the tasks should be complex enough such that they cannot be solved quickly by simple-minded strategies such as random weight guessing.

Recently we discovered (Schmidhuber & Hochreiter, 1996; Hochreiter & Schmidhuber, 1996, 1997) that many long-time-lag tasks used in previous work can be solved more quickly by simple random weight guessing than by the proposed algorithms. For instance, guessing solved a variant of Bengio and Frasconi's parity problem (1994) much faster[4] than the seven methods tested by Bengio et al. (1994) and Bengio and Frasconi (1994). The same is true for some of Miller and Giles's problems (1993). Of course, this does not mean that guessing is a good algorithm. It just means that some previously used problems are not extremely appropriate to demonstrate the quality of previously proposed algorithms.

All our experiments (except experiment 1) involve long minimal time lags; there are no short-time-lag training exemplars facilitating learning. Solutions to most of our tasks are sparse in weight space. They require either many parameters and inputs or high weight precision, such that random weight guessing becomes infeasible.

We always use online learning (as opposed to batch learning) and logistic sigmoids as activation functions. For experiments 1 and 2, initial weights are chosen in the range $[-0.2, 0.2]$, for the other experiments in $[-0.1, 0.1]$. Training sequences are generated randomly according to the various task

---

[4] Different input representations and different types of noise may lead to worse guessing performance (Yoshua Bengio, personal communication, 1996).

descriptions. In slight deviation from the notation in appendix A.1, each discrete time step of each input sequence involves three processing steps: (1) use current input to set the input units, (2) compute activations of hidden units (including input gates, output gates, memory cells), and (3) compute output unit activations. Except for experiments 1, 2a, and 2b, sequence elements are randomly generated online, and error signals are generated only at sequence ends. Net activations are reset after each processed input sequence.

For comparisons with recurrent nets taught by gradient descent, we give results only for RTRL, except for comparison 2a, which also includes BPTT. Note, however, that untruncated BPTT (see, e.g., Williams & Peng, 1990) computes exactly the same gradient as offline RTRL. With long-time-lag problems, offline RTRL (or BPTT) and the online version of RTRL (no activation resets, online weight changes) lead to almost identical, negative results (as confirmed by additional simulations in Hochreiter, 1991; see also Mozer, 1992). This is because offline RTRL, online RTRL, and full BPTT all suffer badly from exponential error decay.

Our LSTM architectures are selected quite arbitrarily. If nothing is known about the complexity of a given problem, a more systematic approach would be to: start with a very small net consisting of one memory cell. If this does not work, try two cells, and so on. Alternatively, use sequential network construction (e.g., Fahlman, 1991).

Following is an outline of the experiments:

- Experiment 1 focuses on a standard benchmark test for recurrent nets: the embedded Reber grammar. Since it allows for training sequences with short time lags, it is not a long-time-lag problem. We include it because it provides a nice example where LSTM's output gates are truly beneficial, and it is a popular benchmark for recurrent nets that has been used by many authors. We want to include at least one experiment where conventional BPTT and RTRL do not fail completely (LSTM, however, clearly outperforms them). The embedded Reber grammar's minimal time lags represent a border case in the sense that it is still possible to learn to bridge them with conventional algorithms. Only slightly longer minimal time lags would make this almost impossible. The more interesting tasks in our article, however, are those that RTRL, BPTT, and others cannot solve at all.

- Experiment 2 focuses on noise-free and noisy sequences involving numerous input symbols distracting from the few important ones. The most difficult task (task 2c) involves hundreds of distractor symbols at random positions and minimal time lags of 1000 steps. LSTM solves it; BPTT and RTRL already fail in case of 10-step minimal time lags (see also Hochreiter, 1991; Mozer, 1992). For this reason RTRL and BPTT are omitted in the remaining, more complex experiments, all of which involve much longer time lags.

- Experiment 3 addresses long-time-lag problems with noise and signal on the same input line. Experiments 3a and 3b focus on Bengio et al.'s 1994 two-sequence problem. Because this problem can be solved quickly by random weight guessing, we also include a far more difficult two-sequence problem (experiment 3c), which requires learning real-valued, conditional expectations of noisy targets, given the inputs.

- Experiments 4 and 5 involve distributed, continuous-valued input representations and require learning to store precise, real values for very long time periods. Relevant input signals can occur at quite different positions in input sequences. Again minimal time lags involve hundreds of steps. Similar tasks never have been solved by other recurrent net algorithms.

- Experiment 6 involves tasks of a different complex type that also has not been solved by other recurrent net algorithms. Again, relevant input signals can occur at quite different positions in input sequences. The experiment shows that LSTM can extract information conveyed by the temporal order of widely separated inputs.

Section 5.7 provides a detailed summary of experimental conditions in two tables for reference.

5.1 Experiment 1: Embedded Reber Grammar.

5.1.1 Task. Our first task is to learn the embedded Reber grammar (Smith & Zipser, 1989; Cleeremans, Servan-Schreiber, & McClelland, 1989; Fahlman,1991). Since it allows for training sequences with short time lags (of as few as nine steps), it is not a long-time-lag problem. We include it for two reasons: (1) it is a popular recurrent net benchmark used by many authors, and we wanted to have at least one experiment where RTRL and BPTT do not fail completely, and (2) it shows nicely how output gates can be beneficial.

Starting at the left-most node of the directed graph in Figure 3, symbol strings are generated sequentially (beginning with the empty string) by following edges—and appending the associated symbols to the current string—until the right-most node is reached (the Reber grammar substrings are analogously generated from Figure 4). Edges are chosen randomly if there is a choice (probability: 0.5). The net's task is to read strings, one symbol at a time, and to predict the next symbol (error signals occur at every time step). To predict the symbol before last, the net has to remember the second symbol.

5.1.2 Comparison. We compare LSTM to Elman nets trained by Elman's training procedure (ELM) (results taken from Cleeremans et al., 1989), Fahlman's recurrent cascade-correlation (RCC) (results taken from Fahlman,
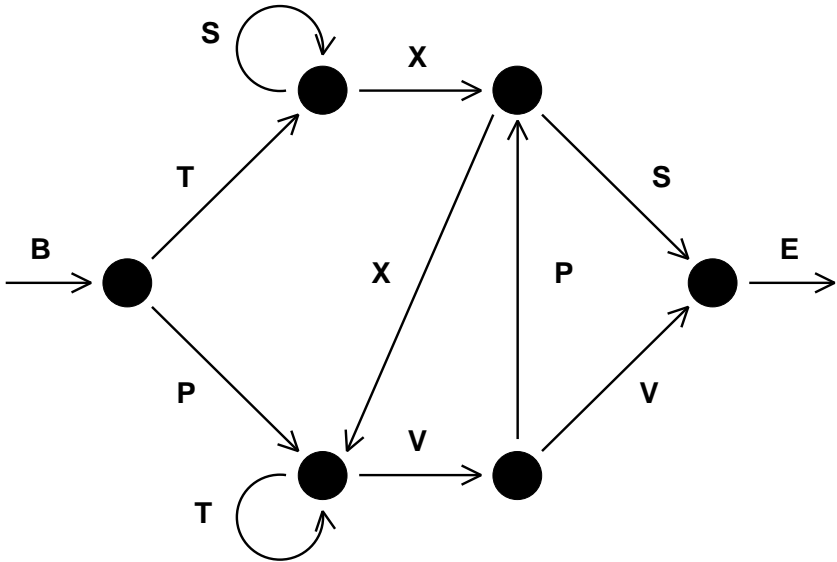
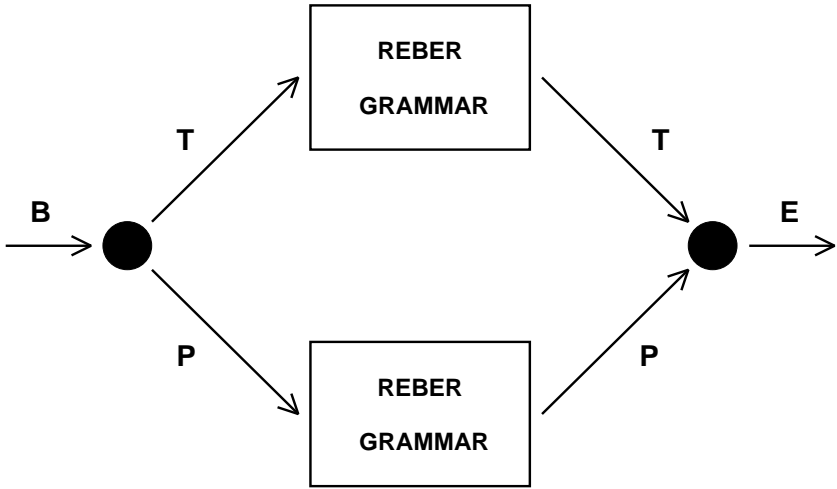Figure 3: Transition diagram for the Reber grammar.

Figure 4: Transition diagram for the embedded Reber grammar. Each box represents a copy of the Reber grammar (see Figure 3).

1991), and RTRL (results taken from Smith & Zipser, 1989), where only the few successful trials are listed). Smith and Zipser actually make the task easier by increasing the probability of short-time-lag exemplars. We did not do this for LSTM.

5.1.3 Training/Testing.   We use a local input-output representation (seven input units, seven output units). Following Fahlman, we use 256 training strings and 256 separate test strings. The training set is generated randomly; training exemplars are picked randomly from the training set. Test sequences are generated randomly, too, but sequences already used in the training set are not used for testing. After string presentation, all activations are reinitialized with zeros. A trial is considered successful if all string symbols of all sequences in both test set and training set are predicted correctly—that is, if the output unit(s) corresponding to the possible next symbol(s) is(are) always the most active ones.

5.1.4 Architectures.   Architectures for RTRL, ELM, and RCC are reported in the references listed above. For LSTM, we use three (four) memory cell blocks. Each block has two (one) memory cells. The output layer's only incoming connections originate at memory cells. Each memory cell and each gate unit receives incoming connections from all memory cells and gate units (the hidden layer is fully connected; less connectivity may work as well). The input layer has forward connections to all units in the hidden layer. The gate units are biased. These architecture parameters make it easy to store at least three input signals (architectures 3-2 and 4-1 are employed to obtain comparable numbers of weights for both architectures: 264 for 4-1 and 276 for 3-2). Other parameters may be appropriate as well, however. All sigmoid functions are logistic with output range $[0, 1]$, except for h, whose range is $[-1, 1]$, and g, whose range is $[-2, 2]$. All weights are initialized in $[-0.2, 0.2]$, except for the output gate biases, which are initialized to -1, -2, and -3, respectively (see abuse problem, solution 2 of section 4). We tried learning rates of 0.1, 0.2, and 0.5.

5.1.5 Results.   We use three different, randomly generated pairs of training and test sets. With each such pair we run 10 trials with different initial weights. See Table 1 for results (mean of 30 trials). Unlike the other methods, LSTM always learns to solve the task. Even when we ignore the unsuccessful trials of the other approaches, LSTM learns much faster.

5.1.6 Importance of Output Gates.   The experiment provides a nice example where the output gate is truly beneficial. Learning to store the first T or P should not perturb activations representing the more easily learnable transitions of the original Reber grammar. This is the job of the output gates. Without output gates, we did not achieve fast learning.

Table 1: Experiment 1: Embedded Reber Grammar.

Method	Hidden Units	Number of Weights	Learning Rate	% of Success	After
RTRL	3	≈ 170	0.05	Some fraction	173,000
RTRL	12	≈ 494	0.1	Some fraction	25,000
ELM	15	≈ 435		0	>200,000
RCC	7–9	≈ 119–198		50	182,000
LSTM	4 blocks, size 1	264	0.1	100	39,740
LSTM	3 blocks, size 2	276	0.1	100	21,730
LSTM	3 blocks, size 2	276	0.2	97	14,060
LSTM	4 blocks, size 1	264	0.5	97	9500
LSTM	3 blocks, size 2	276	0.5	100	8440

Notes: Percentage of successful trials and number of sequence presentations until success for RTRL (results taken from Smith & Zipser, 1989), Elman net trained by Elman's procedure (results taken from Cleeremans et al., 1989), recurrent cascade-correlation (results taken from Fahlman, 1991), and our new approach (LSTM). Weight numbers in the first four rows are estimates, the corresponding papers do not provide all the technical details. Only LSTM almost always learns to solve the task (only 2 failures out of 150 trials). Even when we ignore the unsuccessful trials of the other approaches, LSTM learns much faster (the number of required training examples in the bottom row varies between 3800 and 24,100).

5.2 Experiment 2: Noise-Free and Noisy Sequences.

5.2.1 Task 2a: Noise-Free Sequences with Long Time Lags. There are $p + 1$ possible input symbols denoted $a_1, \ldots, a_{p-1}, a_p = x, a_{p+1} = y$. a_i is locally represented by the $p + 1$-dimensional vector whose ith component is 1 (all other components are 0). A net with $p + 1$ input units and $p + 1$ output units sequentially observes input symbol sequences, one at a time, permanently trying to predict the next symbol; error signals occur at every time step. To emphasize the long-time-lag problem, we use a training set consisting of only two very similar sequences: $(y, a_1, a_2, \ldots, a_{p-1}, y)$ and $(x, a_1, a_2, \ldots, a_{p-1}, x)$. Each is selected with probability 0.5. To predict the final element, the net has to learn to store a representation of the first element for p time steps.

We compare real-time recurrent learning for fully recurrent nets (RTRL), back-propagation through time (BPTT), the sometimes very successful two-net neural sequence chunker (CH; Schmidhuber, 1992b), and our new method (LSTM). In all cases, weights are initialized in $[-0.2, 0.2]$. Due to limited computation time, training is stopped after 5 million sequence presentations. A successful run is one that fulfills the following criterion: after training, during 10,000 successive, randomly chosen input sequences, the maximal absolute error of all output units is always below 0.25.

Table 2: Task 2a: Percentage of Successful Trials and Number of Training Sequences until Success.

Method	Delay p	Learning Rate	Number of Weights	% Successful Trials	Success After
RTRL	4	1.0	36	78	1,043,000
RTRL	4	4.0	36	56	892,000
RTRL	4	10.0	36	22	254,000
RTRL	10	1.0–10.0	144	0	> 5,000,000
RTRL	100	1.0–10.0	10404	0	> 5,000,000
BPTT	100	1.0–10.0	10404	0	> 5,000,000
CH	100	1.0	10506	33	32,400
LSTM	100	1.0	10504	100	5,040

Notes: Table entries refer to means of 18 trials. With 100 time-step delays, only CH and LSTM achieve successful trials. Even when we ignore the unsuccessful trials of the other approaches, LSTM learns much faster.

Architectures.

RTRL: One self-recurrent hidden unit, $p + 1$ nonrecurrent output units. Each layer has connections from all layers below. All units use the logistic activation function sigmoid in [0, 1].

BPTT: Same architecture as the one trained by RTRL.

CH: Both net architectures like RTRL's, but one has an additional output for predicting the hidden unit of the other one (see Schmidhuber, 1992b, for details).

LSTM: As with RTRL, but the hidden unit is replaced by a memory cell and an input gate (no output gate required). g is the logistic sigmoid, and h is the identity function $h : h(x) = x, \forall x$. Memory cell and input gate are added once the error has stopped decreasing (see abuse problem: solution 1 in section 4).

Results. Using RTRL and a short four-time-step delay ($p = 4$), 7/9 of all trials were successful. No trial was successful with $p = 10$. With long time lags, only the neural sequence chunker and LSTM achieved successful trials; BPTT and RTRL failed. With $p = 100$, the two-net sequence chunker solved the task in only one-third of all trials. LSTM, however, always learned to solve the task. Comparing successful trials only, LSTM learned much faster. See Table 2 for details. It should be mentioned, however, that a hierarchical chunker can also always quickly solve this task (Schmidhuber, 1992c, 1993).

5.2.2 Task 2b: No Local Regularities. With task 2a, the chunker sometimes learns to predict the final element correctly, but only because of pre-

dictable local regularities in the input stream that allow for compressing the sequence. In a more difficult task, involving many more different possible sequences, we remove compressibility by replacing the deterministic subsequence $(a_1, a_2, \ldots, a_{p-1})$ by a random subsequence (of length $p-1$) over the alphabet $a_1, a_2, \ldots, a_{p-1}$. We obtain two classes (two sets of sequences) $\{(y, a_{i_1}, a_{i_2}, \ldots, a_{i_{p-1}}, y) \mid 1 \leq i_1, i_2, \ldots, i_{p-1} \leq p-1\}$ and $\{(x, a_{i_1}, a_{i_2}, \ldots, a_{i_{p-1}}, x) \mid 1 \leq i_1, i_2, \ldots, i_{p-1} \leq p-1\}$. Again, every next sequence element has to be predicted. The only totally predictable targets, however, are x and y, which occur at sequence ends. Training exemplars are chosen randomly from the two classes. Architectures and parameters are the same as in experiment 2a. A successful run is one that fulfills the following criterion: after training, during 10,000 successive, randomly chosen input sequences, the maximal absolute error of all output units is below 0.25 at sequence end.

Results.   As expected, the chunker failed to solve this task (so did BPTT and RTRL, of course). LSTM, however, was always successful. On average (mean of 18 trials), success for $p = 100$ was achieved after 5680 sequence presentations. This demonstrates that LSTM does not require sequence regularities to work well.

5.2.3 Task 2c: Very Long Time Lags—No Local Regularities.   This is the most difficult task in this subsection. To our knowledge, no other recurrent net algorithm can solve it. Now there are $p + 4$ possible input symbols denoted $a_1, \ldots, a_{p-1}, a_p, a_{p+1} = e, a_{p+2} = b, a_{p+3} = x, a_{p+4} = y$. $a_1, \ldots, a_p$ are also called distractor symbols. Again, a_i is locally represented by the $p + 4$-dimensional vector whose ith component is 1 (all other components are 0). A net with $p + 4$ input units and 2 output units sequentially observes input symbol sequences, one at a time. Training sequences are randomly chosen from the union of two very similar subsets of sequences: $\{(b, y, a_{i_1}, a_{i_2}, \ldots, a_{i_{q+k}}, e, y) \mid 1 \leq i_1, i_2, \ldots, i_{q+k} \leq q\}$ and $\{(b, x, a_{i_1}, a_{i_2}, \ldots, a_{i_{q+k}}, e, x) \mid 1 \leq i_1, i_2, \ldots, i_{q+k} \leq q\}$. To produce a training sequence, we randomly generate a sequence prefix of length $q + 2$, randomly generate a sequence suffix of additional elements ($\neq b, e, x, y$) with probability 9/10 or, alternatively, an e with probability 1/10. In the latter case, we conclude the sequence with x or y, depending on the second element. For a given k, this leads to a uniform distribution on the possible sequences with length $q + k + 4$. The minimal sequence length is $q + 4$; the expected length is

$$4 + \sum_{k=0}^{\infty} \frac{1}{10} \left(\frac{9}{10} \right)^k (q + k) = q + 14.$$

The expected number of occurrences of element a_i, $1 \leq i \leq p$, in a sequence is $(q + 10)/p \approx \frac{q}{p}$. The goal is to predict the last symbol, which always occurs after the "trigger symbol" e. Error signals are generated only at sequence

Table 3: Task 2c: LSTM with Very Long Minimal Time Lags $q + 1$ and a Lot of Noise.

q (Time Lag -1)	p (Number of Random Inputs)	$\frac{q}{p}$	Number of Weights	Success After
50	50	1	364	30,000
100	100	1	664	31,000
200	200	1	1264	33,000
500	500	1	3064	38,000
1000	1,000	1	6064	49,000
1000	500	2	3064	49,000
1000	200	5	1264	75,000
1000	100	10	664	135,000
1000	50	20	364	203,000

Notes: p is the number of available distractor symbols ($p+4$ is the number of input units). q/p is the expected number of occurrences of a given distractor symbol in a sequence. The right-most column lists the number of training sequences required by LSTM (BPTT, RTRL, and the other competitors have no chance of solving this task). If we let the number of distractor symbols (and weights) increase in proportion to the time lag, learning time increases very slowly. The lower block illustrates the expected slowdown due to increased frequency of distractor symbols.

ends. To predict the final element, the net has to learn to store a representation of the second element for at least $q + 1$ time steps (until it sees the trigger symbol e). Success is defined as prediction error (for final sequence element) of both output units always below 0.2, for 10,000 successive, randomly chosen input sequences.

Architecture/Learning.   The net has $p + 4$ input units and 2 output units. Weights are initialized in $[-0.2, 0.2]$. To avoid too much learning time variance due to different weight initializations, the hidden layer gets two memory cells (two cell blocks of size 1, although one would be sufficient). There are no other hidden units. The output layer receives connections only from memory cells. Memory cells and gate units receive connections from input units, memory cells, and gate units (the hidden layer is fully connected). No bias weights are used. h and g are logistic sigmoids with output ranges $[-1, 1]$ and $[-2, 2]$, respectively. The learning rate is 0.01. Note that the minimal time lag is $q + 1$; the net never sees short training sequences facilitating the classification of long test sequences.

Results.   Twenty trials were made for all tested pairs (p, q). Table 3 lists the mean of the number of training sequences required by LSTM to achieve success (BPTT and RTRL have no chance of solving nontrivial tasks with minimal time lags of 1000 steps).

Scaling.   Table 3 shows that if we let the number of input symbols (and weights) increase in proportion to the time lag, learning time increases very slowly. This is another remarkable property of LSTM not shared by any other method we are aware of. Indeed, RTRL and BPTT are far from scaling reasonably; instead, they appear to scale exponentially and appear quite useless when the time lags exceed as few as 10 steps.

Distractor Influence.   In Table 3, the column headed by q/p gives the expected frequency of distractor symbols. Increasing this frequency decreases learning speed, an effect due to weight oscillations caused by frequently observed input symbols.

5.3 Experiment 3: Noise and Signal on Same Channel. This experiment serves to illustrate that LSTM does not encounter fundamental problems if noise and signal are mixed on the same input line. We initially focus on Bengio et al.'s simple 1994 two-sequence problem. In experiment 3c we pose a more challenging two-sequence problem.

5.3.1 Task 3a (Two-Sequence Problem).   The task is to observe and then classify input sequences. There are two classes, each occurring with probability 0.5. There is only one input line. Only the first N real-valued sequence elements convey relevant information about the class. Sequence elements at positions $t > N$ are generated by a gaussian with mean zero and variance 0.2. Case $N = 1$: the first sequence element is 1.0 for class 1, and -1.0 for class 2. Case $N = 3$: the first three elements are 1.0 for class 1 and -1.0 for class 2. The target at the sequence end is 1.0 for class 1 and 0.0 for class 2. Correct classification is defined as absolute output error at sequence end below 0.2. Given a constant T, the sequence length is randomly selected between T and $T + T/10$ (a difference to Bengio et al.'s problem is that they also permit shorter sequences of length $T/2$).

Guessing.   Bengio et al. (1994) and Bengio and Frasconi (1994) tested seven different methods on the two-sequence problem. We discovered, however, that random weight guessing easily outperforms them all because the problem is so simple.[5] See Schmidhuber and Hochreiter (1996) and Hochreiter and Schmidhuber (1996, 1997) for additional results in this vein.

LSTM Architecture.   We use a three-layer net with one input unit, one output unit, and three cell blocks of size 1. The output layer receives connections only from memory cells. Memory cells and gate units receive inputs from input units, memory cells, and gate units and have bias weights. Gate

---

[5] However, different input representations and different types of noise may lead to worse guessing performance (Yoshua Bengio, personal communication, 1996).

Table 4: Task 3a: Bengio et al.'s Two-Sequence Problem.

T	N	Stop: ST1	Stop: ST2	Number of Weights	ST2: Fraction Misclassified
100	3	27,380	39,850	102	0.000195
100	1	58,370	64,330	102	0.000117
1000	3	446,850	452,460	102	0.000078

Notes: *T* is minimal sequence length. *N* is the number of information-conveying elements at sequence begin. The column headed by ST1 (ST2) gives the number of sequence presentations required to achieve stopping criterion ST1 (ST2). The right-most column lists the fraction of misclassified posttraining sequences (with absolute error > 0.2) from a test set consisting of 2560 sequences (tested after ST2 was achieved). All values are means of 10 trials. We discovered, however, that this problem is so simple that random weight guessing solves it faster than LSTM and any other method for which there are published results.

units and output unit are logistic sigmoid in $[0, 1]$, h in $[-1, 1]$, and g in $[-2, 2]$.

Training/Testing.   All weights (except the bias weights to gate units) are randomly initialized in the range $[-0.1, 0.1]$. The first input gate bias is initialized with -1.0, the second with -3.0, and the third with -5.0. The first output gate bias is initialized with -2.0, the second with -4.0, and the third with -6.0. The precise initialization values hardly matter though, as confirmed by additional experiments. The learning rate is 1.0. All activations are reset to zero at the beginning of a new sequence.

We stop training (and judge the task as being solved) according to the following criteria: ST1: none of 256 sequences from a randomly chosen test set is misclassified; ST2: ST1 is satisfied, and mean absolute test set error is below 0.01. In case of ST2, an additional test set consisting of 2560 randomly chosen sequences is used to determine the fraction of misclassified sequences.

Results.   See Table 4. The results are means of 10 trials with different weight initializations in the range $[-0.1, 0.1]$. LSTM is able to solve this problem, though by far not as fast as random weight guessing (see "Guessing" above). Clearly this trivial problem does not provide a very good testbed to compare performance of various nontrivial algorithms. Still, it demonstrates that LSTM does not encounter fundamental problems when faced with signal and noise on the same channel.

5.3.2 Task 3b.   The architecture, parameters, and other elements are as in task 3a, but now with gaussian noise (mean 0 and variance 0.2) added to the

Table 5: Task 3b: Modified Two-Sequence Problem.

T	N	Stop: ST1	Stop: ST2	Number of Weights	ST2: Fraction Misclassified
100	3	41,740	43,250	102	0.00828
100	1	74,950	78,430	102	0.01500
1000	1	481,060	485,080	102	0.01207

Note: Same as in Table 4, but now the information-conveying elements are also perturbed by noise.

information-conveying elements ($t <= N$). We stop training (and judge the task as being solved) according to the following, slightly redefined criteria: ST1: fewer than 6 out of 256 sequences from a randomly chosen test set are misclassified; ST2: ST1 is satisfied, and mean absolute test set error is below 0.04. In case of ST2, an additional test set consisting of 2560 randomly chosen sequences is used to determine the fraction of misclassified sequences.

Results.   See Table 5. The results represent means of 10 trials with different weight initializations. LSTM easily solves the problem.

5.3.3 Task 3c.   The architecture, parameters, and other elements are as in task 3a, but with a few essential changes that make the task nontrivial: the targets are 0.2 and 0.8 for class 1 and class 2, respectively, and there is gaussian noise on the targets (mean 0 and variance 0.1; S.D. 0.32). To minimize mean squared error, the system has to learn the conditional expectations of the targets given the inputs. Misclassification is defined as absolute difference between output and noise-free target (0.2 for class 1 and 0.8 for class 2) > 0.1. The network output is considered acceptable if the mean absolute difference between noise-free target and output is below 0.015. Since this requires high weight precision, task 3c (unlike tasks 3a and 3b) cannot be solved quickly by random guessing.

Training/Testing.   The learning rate is 0.1. We stop training according to the following criterion: none of 256 sequences from a randomly chosen test set is misclassified, and mean absolute difference between the noise-free target and output is below 0.015. An additional test set consisting of 2560 randomly chosen sequences is used to determine the fraction of misclassified sequences.

Results.   See Table 6. The results represent means of 10 trials with different weight initializations. Despite the noisy targets, LSTM still can solve the problem by learning the expected target values.

Table 6: Task 3c: Modified, More Challenging Two-Sequence Problem.

T	N	Stop	Number of Weights	Fraction Misclassified	Average Difference to Mean
100	3	269,650	102	0.00558	0.014
100	1	565,640	102	0.00441	0.012

Notes: Same as in Table 4, but with noisy real-valued targets. The system has to learn the conditional expectations of the targets given the inputs. The right-most column provides the average difference between network output and expected target. Unlike tasks 3a and 3b, this one cannot be solved quickly by random weight guessing.

5.4 Experiment 4: Adding Problem. The difficult task in this section is of a type that has never been solved by other recurrent net algorithms. It shows that LSTM can solve long-time-lag problems involving distributed, continuous-valued representations.

5.4.1 Task. Each element of each input sequence is a pair of components. The first component is a real value randomly chosen from the interval $[-1, 1]$; the second is 1.0, 0.0, or -1.0 and is used as a marker. At the end of each sequence, the task is to output the sum of the first components of those pairs that are marked by second components equal to 1.0. Sequences have random lengths between the minimal sequence length T and $T + T/10$. In a given sequence, exactly two pairs are marked, as follows: we first randomly select and mark one of the first 10 pairs (whose first component we call X_1). Then we randomly select and mark one of the first $T/2 - 1$ still unmarked pairs (whose first component we call X_2). The second components of all remaining pairs are zero except for the first and final pair, whose second components are -1. (In the rare case where the first pair of the sequence gets marked, we set X_1 to zero.) An error signal is generated only at the sequence end: the target is $0.5 + (X_1 + X_2)/4.0$ (the sum $X_1 + X_2$ scaled to the interval $[0, 1]$). A sequence is processed correctly if the absolute error at the sequence end is below 0.04.

5.4.2 Architecture. We use a three-layer net with two input units, one output unit, and two cell blocks of size 2. The output layer receives connections only from memory cells. Memory cells and gate units receive inputs from memory cells and gate units (the hidden layer is fully connected; less connectivity may work as well). The input layer has forward connections to all units in the hidden layer. All noninput units have bias weights. These architecture parameters make it easy to store at least two input signals (a cell block size of 1 works well, too). All activation functions are logistic with output range $[0, 1]$, except for h, whose range is $[-1, 1]$, and g, whose range is $[-2, 2]$.

Table 7: Experiment 4: Results for the Adding Problem.

T	Minimal Lag	Number of Weights	Number of Wrong Predictions	Success After
100	50	93	1 out of 2560	74,000
500	250	93	0 out of 2560	209,000
1000	500	93	1 out of 2560	853,000

Notes: T is the minimal sequence length, $T/2$ the minimal time lag. "Number of Wrong Predictions" is the number of incorrectly processed sequences (error > 0.04) from a test set containing 2560 sequences. The right-most column gives the number of training sequences required to achieve the stopping criterion. All values are means of 10 trials. For $T = 1000$ the number of required training examples varies between 370,000 and 2,020,000, exceeding 700,000 in only three cases.

5.4.3 State Drift Versus Initial Bias. Note that the task requires storing the precise values of real numbers for long durations; the system must learn to protect memory cell contents against even minor internal state drift (see section 4). To study the significance of the drift problem, we make the task even more difficult by biasing all noninput units, thus artificially inducing internal state drift. All weights (including the bias weights) are randomly initialized in the range $[-0.1, 0.1]$. Following section 4's remedy for state drifts, the first input gate bias is initialized with -3.0 and the second with -6.0 (though the precise values hardly matter, as confirmed by additional experiments).

5.4.4 Training/Testing. The learning rate is 0.5. Training is stopped once the average training error is below 0.01, and the 2000 most recent sequences were processed correctly.

5.4.5 Results. With a test set consisting of 2560 randomly chosen sequences, the average test set error was always below 0.01, and there were never more than three incorrectly processed sequences. Table 7 shows details.

The experiment demonstrates that LSTM is able to work well with distributed representations, LSTM is able to learn to perform calculations involving continuous values, and since the system manages to store continuous values without deterioration for minimal delays of $T/2$ time steps, there is no significant, harmful internal state drift.

5.5 Experiment 5: Multiplication Problem. One may argue that LSTM is a bit biased toward tasks such as the adding problem from the previous subsection. Solutions to the adding problem may exploit the CEC's built-in integration capabilities. Although this CEC property may be viewed as a

Table 8: Experiment 5: Results for the Multiplication Problem.

T	Minimal Lag	Number of Weights	n_{seq}	Number of Wrong Predictions	MSE	Success After
100	50	93	140	139 out of 2560	0.0223	482,000
100	50	93	13	14 out of 2560	0.0139	1,273,000

Notes: T is the minimal sequence length and $T/2$ the minimal time lag. We test on a test set containing 2560 sequences as soon as less than n_{seq} of the 2000 most recent training sequences lead to error > 0.04. "Number of Wrong Predictions" is the number of test sequences with error > 0.04. MSE is the mean squared error on the test set. The right-most column lists numbers of training sequences required to achieve the stopping criterion. All values are means of 10 trials.

feature rather than a disadvantage (integration seems to be a natural subtask of many tasks occurring in the real world), the question arises whether LSTM can also solve tasks with inherently nonintegrative solutions. To test this, we change the problem by requiring the final target to equal the product (instead of the sum) of earlier marked inputs.

5.5.1 Task. This is like the task in section 5.4, except that the first component of each pair is a real value randomly chosen from the interval [0, 1]. In the rare case where the first pair of the input sequence gets marked, we set X_1 to 1.0. The target at sequence end is the product $X_1 \times X_2$.

5.5.2 Architecture. This is as in section 5.4. All weights (including the bias weights) are randomly initialized in the range $[-0.1, 0.1]$.

5.5.3 Training/Testing. The learning rate is 0.1. We test performance twice: as soon as less than n_{seq} of the 2000 most recent training sequences lead to absolute errors exceeding 0.04, where $n_{seq} = 140$ and $n_{seq} = 13$. Why these values? $n_{seq} = 140$ is sufficient to learn storage of the relevant inputs. It is not enough though to fine-tune the precise final outputs. $n_{seq} = 13$, however, leads to quite satisfactory results.

5.5.4 Results. For $n_{seq} = 140$ ($n_{seq} = 13$) with a test set consisting of 2560 randomly chosen sequences, the average test set error was always below 0.026 (0.013), and there were never more than 170 (15) incorrectly processed sequences. Table 8 shows details. (A net with additional standard hidden units or with a hidden layer above the memory cells may learn the fine-tuning part more quickly.)

The experiment demonstrates that LSTM can solve tasks involving both continuous-valued representations and nonintegrative information processing.

5.6 Experiment 6: Temporal Order. In this subsection, LSTM solves other difficult (but artificial) tasks that have never been solved by previous recurrent net algorithms. The experiment shows that LSTM is able to extract information conveyed by the temporal order of widely separated inputs.

5.6.1 Task 6a: Two Relevant, Widely Separated Symbols. The goal is to classify sequences. Elements and targets are represented locally (input vectors with only one nonzero bit). The sequence starts with an *E*, ends with a *B* (the "trigger symbol"), and otherwise consists of randomly chosen symbols from the set $\{a, b, c, d\}$ except for two elements at positions t_1 and t_2 that are either *X* or *Y*. The sequence length is randomly chosen between 100 and 110, t_1 is randomly chosen between 10 and 20, and t_2 is randomly chosen between 50 and 60. There are four sequence classes *Q*, *R*, *S*, *U*, which depend on the temporal order of *X* and *Y*. The rules are: $X, X \to Q$; $X, Y \to R$; $Y, X \to S$; $Y, Y \to U$.

5.6.2 Task 6b: Three Relevant, Widely Separated Symbols. Again, the goal is to classify sequences. Elements and targets are represented locally. The sequence starts with an *E*, ends with a *B* (the trigger symbol), and otherwise consists of randomly chosen symbols from the set $\{a, b, c, d\}$ except for three elements at positions t_1, t_2, and t_3 that are either *X* or *Y*. The sequence length is randomly chosen between 100 and 110, t_1 is randomly chosen between 10 and 20, t_2 is randomly chosen between 33 and 43, and t_3 is randomly chosen between 66 and 76. There are eight sequence classes— *Q*, *R*, *S*, *U*, *V*, *A*, *B*, *C*—which depend on the temporal order of the *X*s and *Y*s. The rules are: $X, X, X \to Q$; $X, X, Y \to R$; $X, Y, X \to S$; $X, Y, Y \to U$; $Y, X, X \to V$; $Y, X, Y \to A$; $Y, Y, X \to B$; $Y, Y, Y \to C$.

There are as many output units as there are classes. Each class is locally represented by a binary target vector with one nonzero component. With both tasks, error signals occur only at the end of a sequence. The sequence is classified correctly if the final absolute error of all output units is below 0.3.

Architecture. We use a three-layer net with eight input units, two (three) cell blocks of size 2, and four (eight) output units for task 6a (6b). Again all noninput units have bias weights, and the output layer receives connections from memory cells, only. Memory cells and gate units receive inputs from input units, memory cells, and gate units (the hidden layer is fully connected; less connectivity may work as well). The architecture parameters for task 6a (6b) make it easy to store at least two (three) input signals. All activation functions are logistic with output range $[0, 1]$, except for *h*, whose range is $[-1, 1]$, and *g*, whose range is $[-2, 2]$.

Table 9: Experiment 6: Results for the Temporal Order Problem.

Task	Number of Weights	Number of Wrong Predictions	Success After
Task 6a	156	1 out of 2560	31,390
Task 6b	308	2 out of 2560	571,100

Notes: "Number of Wrong Predictions" is the number of incorrectly classified sequences (error > 0.3 for at least one output unit) from a test set containing 2560 sequences. The right-most column gives the number of training sequences required to achieve the stopping criterion. The results for task 6a are means of 20 trials; those for task 6b of 10 trials.

Training/Testing. The learning rate is 0.5 (0.1) for experiment 6a (6b). Training is stopped once the average training error falls below 0.1 and the 2000 most recent sequences were classified correctly. All weights are initialized in the range $[-0.1, 0.1]$. The first input gate bias is initialized with -2.0, the second with -4.0, and (for experiment 6b) the third with -6.0 (again, we confirmed by additional experiments that the precise values hardly matter).

Results. With a test set consisting of 2560 randomly chosen sequences, the average test set error was always below 0.1, and there were never more than three incorrectly classified sequences. Table 9 shows details.

The experiment shows that LSTM is able to extract information conveyed by the temporal order of widely separated inputs. In task 6a, for instance, the delays between the first and second relevant input and between the second relevant input and sequence end are at least 30 time steps.

Typical Solutions. In experiment 6a, how does LSTM distinguish between temporal orders (X, Y) and (Y, X)? One of many possible solutions is to store the first X or Y in cell block 1 and the second X/Y in cell block 2. Before the first X/Y occurs, block 1 can see that it is still empty by means of its recurrent connections. After the first X/Y, block 1 can close its input gate. Once block 1 is filled and closed, this fact will become visible to block 2 (recall that all gate units and all memory cells receive connections from all nonoutput units).

Typical solutions, however, require only one memory cell block. The block stores the first X or Y; once the second X/Y occurs, it changes its state depending on the first stored symbol. Solution type 1 exploits the connection between memory cell output and input gate unit. The following events cause different input gate activations: X occurs in conjunction with a filled block; X occurs in conjunction with an empty block. Solution type 2 is based on a strong, positive connection between memory cell output and memory cell input. The previous occurrence of X (Y) is represented by a

Table 10: Summary of Experimental Conditions for LSTM, Part I.

(1) Task	(2) p	(3) lag	(4) b	(5) s	(6) in	(7) out	(8) w	(9) c	(10) ogb	(11) igb	(12) bias	(13) h	(14) g	(15) α
1-1	9	9	4	1	7	7	264	F	−1, −2, −3, −4	r	ga	h1	g2	0.1
1-2	9	9	3	2	7	7	276	F	−1, −2, −3	r	ga	h1	g2	0.1
1-3	9	9	3	2	7	7	276	F	−1, −2, −3	r	ga	h1	g2	0.2
1-4	9	9	4	1	7	7	264	F	−1, −2, −3, −4	r	ga	h1	g2	0.5
1-5	9	9	3	2	7	7	276	F	−1, −2, −3	r	ga	h1	g2	0.5
2a	100	100	1	1	101	101	10,504	B	No og	None	None	id	g1	1.0
2b	100	100	1	1	101	101	10,504	B	No og	None	None	id	g1	1.0
2c-1	50	50	2	1	54	2	364	F	None	None	None	h1	g2	0.01
2c-2	100	100	2	1	104	2	664	F	None	None	None	h1	g2	0.01
2c-3	200	200	2	1	204	2	1264	F	None	None	None	h1	g2	0.01
2c-4	500	500	2	1	504	2	3064	F	None	None	None	h1	g2	0.01
2c-5	1000	1000	2	1	1004	2	6064	F	None	None	None	h1	g2	0.01
2c-6	1000	1000	2	1	504	2	3064	F	None	None	None	h1	g2	0.01
2c-7	1000	1000	2	1	204	2	1264	F	None	None	None	h1	g2	0.01
2c-8	1000	1000	2	1	104	2	664	F	None	None	None	h1	g2	0.01
2c-9	1000	1000	2	1	54	2	364	F	None	None	None	h1	g2	0.01
3a	100	100	3	1	1	1	102	F	−2, −4, −6	−1, −3, −5	b1	h1	g2	1.0
3b	100	100	3	1	1	1	102	F	−2, −4, −6	−1, −3, −5	b1	h1	g2	1.0
3c	100	100	3	1	1	1	102	F	−2, −4, −6	−1, −3, −5	b1	h1	g2	0.1
4-1	100	50	2	2	2	1	93	F	r	−3, −6	All	h1	g2	0.5
4-2	500	250	2	2	2	1	93	F	r	−3, −6	All	h1	g2	0.5
4-3	1000	500	2	2	2	1	93	F	r	−3, −6	All	h1	g2	0.5
5	100	50	2	2	2	1	93	F	r	r	All	h1	g2	0.1
6a	100	40	2	2	8	4	156	F	r	−2, −4	All	h1	g2	0.5
6b	100	24	3	2	8	8	308	F	r	−2, −4, −6	All	h1	g2	0.1

Notes: Col. 1: task number. Col. 2: minimal sequence length p. Col. 3: minimal number of steps between most recent relevant input information and teacher signal. Col. 4: number of cell blocks b. Col. 5: block size s. Col. 6: Number of input units in. Col. 7: Number of output units out. Col. 8: number of weights w. Col. 9: c describes connectivity: F means "output layer receives connections from memory cells; memory cells and gate units receive connections from input units, memory cells and gate units"; B means "each layer receives connections from all layers below." Col. 10: Initial output gate bias ogb, where r stands for "randomly chosen from the interval $[-0.1, 0.1]$" and no og means "no output gate used." Col. 11: initial input gate bias igb (see Col. 10). Col. 12: which units have bias weights? $b1$ stands for "all hidden units", ga for "only gate units," and all for "all noninput units." Col. 13: the function h, where id is identity function, $h1$ is logistic sigmoid in $[-2, 2]$. Col. 14: the logistic function g, where $g1$ is sigmoid in $[0, 1]$, $g2$ in $[-1, 1]$. Col. 15: learning rate α.

positive (negative) internal state. Once the input gate opens for the second time, so does the output gate, and the memory cell output is fed back to its own input. This causes (X, Y) to be represented by a positive internal state, because X contributes to the new internal state twice (via current internal state and cell output feedback). Similarly, (Y, X) gets represented by a negative internal state.

5.7 Summary of Experimental Conditions. Tables 10 and 11 provide an overview of the most important LSTM parameters and architectural details for experiments 1 through 6. The conditions of the simple experiments 2a

Table 11: Summary of Experimental Conditions for LSTM, Part II.

(1) Task	(2) Select	(3) Interval	(4) Test Set Size	(5) Stopping Criterion	(6) Success
1	t1	$[-0.2, 0.2]$	256	Training and test correctly pred.	See text
2a	t1	$[-0.2, 0.2]$	no test set	After 5 million exemplars	ABS(0.25)
2b	t2	$[-0.2, 0.2]$	10,000	After 5 million exemplars	ABS(0.25)
2c	t2	$[-0.2, 0.2]$	10,000	After 5 million exemplars	ABS(0.2)
3a	t3	$[-0.1, 0.1]$	2560	ST1 and ST2 (see text)	ABS(0.2)
3b	t3	$[-0.1, 0.1]$	2560	ST1 and ST2 (see text)	ABS(0.2)
3c	t3	$[-0.1, 0.1]$	2560	ST1 and ST2 (see text)	See text
4	t3	$[-0.1, 0.1]$	2560	ST3(0.01)	ABS(0.04)
5	t3	$[-0.1, 0.1]$	2560	see text	ABS(0.04)
6a	t3	$[-0.1, 0.1]$	2560	ST3(0.1)	ABS(0.3)
6b	t3	$[-0.1, 0.1]$	2560	ST3(0.1)	ABS(0.3)

Notes: Col. 1: task number. Col. 2: training exemplar selelction, where $t1$ stands for "randomly chosen form training set," $t2$ for "randomly chosen from two classes," and $t3$ for "randomly generated on line." Col. 3: weight initialization interval. Col. 4: test set size. Col. 5: Stopping criterion for training, where $ST3(\beta)$ stands for "average training error below β and the 2000 most recent sequences were processed correctly." Col. 6: success (correct classification) criterion, where $ABS(\beta)$ stands for "absolute error of all output units at sequence end is below β."

and 2b differ slightly from those of the other, more systematic experiments, due to historical reasons.

6 Discussion

6.1 Limitations of LSTM.

- The particularly efficient truncated backpropagation version of the LSTM algorithm will not easily solve problems similar to strongly delayed XOR problems, where the goal is to compute the XOR of two widely separated inputs that previously occurred somewhere in a noisy sequence. The reason is that storing only one of the inputs will not help to reduce the expected error; the task is nondecomposable in the sense that it is impossible to reduce the error incrementally by first solving an easier subgoal.

  In theory, this limitation can be circumvented by using the full gradient (perhaps with additional conventional hidden units receiving input from the memory cells). But we do not recommend computing the full gradient for the following reasons: (1) It increases computational complexity, (2) constant error flow through CECs can be shown only for truncated LSTM, and (3) we actually did conduct a few experiments with nontruncated LSTM. There was no significant difference to truncated LSTM, exactly because outside the CECs, error flow tends

to vanish quickly. For the same reason, full BPTT does not outperform truncated BPTT.

- Each memory cell block needs two additional units (input and output gate). In comparison to standard recurrent nets, however, this does not increase the number of weights by more than a factor of 9: each conventional hidden unit is replaced by at most three units in the LSTM architecture, increasing the number of weights by a factor of 3^2 in the fully connected case. Note, however, that our experiments use quite comparable weight numbers for the architectures of LSTM and competing approaches.

- Due to its constant error flow through CECs within memory cells, LSTM generally runs into problems similar to those of feedforward nets' seeing the entire input string at once. For instance, there are tasks that can be quickly solved by random weight guessing but not by the truncated LSTM algorithm with small weight initializations, such as the 500-step parity problem (see the introduction to section 5). Here, LSTM's problems are similar to the ones of a feedforward net with 500 inputs, trying to solve 500-bit parity. Indeed LSTM typically behaves much like a feedforward net trained by backpropagation that sees the entire input. But that is also precisely why it so clearly outperforms previous approaches on many nontrivial tasks with significant search spaces.

- LSTM does not have any problems with the notion of recency that go beyond those of other approaches. All gradient-based approaches, however, suffer from a practical inability to count discrete time steps precisely. If it makes a difference whether a certain signal occurred 99 or 100 steps ago, then an additional counting mechanism seems necessary. Easier tasks, however, such as one that requires making a difference only between, say, 3 and 11 steps, do not pose any problems to LSTM. For instance, by generating an appropriate negative connection between memory cell output and input, LSTM can give more weight to recent inputs and learn decays where necessary.

6.2 Advantages of LSTM.

- The constant error backpropagation within memory cells results in LSTM's ability to bridge very long time lags in case of problems similar to those discussed above.

- For long-time-lag problems such as those discussed in this article, LSTM can handle noise, distributed representations, and continuous values. In contrast to finite state automata or hidden Markov models, LSTM does not require an a priori choice of a finite number of states. In principle, it can deal with unlimited state numbers.

- For problems discussed in this article, LSTM generalizes well, even if the positions of widely separated, relevant inputs in the input sequence do not matter. Unlike previous approaches, ours quickly learns to distinguish between two or more widely separated occurrences of a particular element in an input sequence, without depending on appropriate short-time-lag training exemplars.

- There appears to be no need for parameter fine tuning. LSTM works well over a broad range of parameters such as learning rate, input gate bias, and output gate bias. For instance, to some readers the learning rates used in our experiments may seem large. However, a large learning rate pushes the output gates toward zero, thus automatically countermanding its own negative effects.

- The LSTM algorithm's update complexity per weight and time step is essentially that of BPTT, namely, $O(1)$. This is excellent in comparison to other approaches such as RTRL. Unlike full BPTT, however, LSTM is local in both space and time.

7 Conclusion

Each memory cell's internal architecture guarantees constant error flow within its CEC, provided that truncated backpropagation cuts off error flow trying to leak out of memory cells. This represents the basis for bridging very long time lags. Two gate units learn to open and close access to error flow within each memory cell's CEC. The multiplicative input gate affords protection of the CEC from perturbation by irrelevant inputs. Similarly, the multiplicative output gate protects other units from perturbation by currently irrelevant memory contents.

To find out about LSTM's practical limitations we intend to apply it to real-world data. Application areas will include time-series prediction, music composition, and speech processing. It will also be interesting to augment sequence chunkers (Schmidhuber, 1992b, 1993) by LSTM to combine the advantages of both.

Appendix

A.1 Algorithm Details. In what follows, the index k ranges over output units, i ranges over hidden units, c_j stands for the jth memory cell block, c_j^v denotes the vth unit of memory cell block c_j, u, l, m stand for arbitrary units, and t ranges over all time steps of a given input sequence.

The gate unit logistic sigmoid (with range $[0, 1]$) used in the experiments is

$$f(x) = \frac{1}{1 + \exp(-x)} \ . \tag{A.1}$$

The function h (with range $[-1, 1]$) used in the experiments is

$$h(x) = \frac{2}{1 + \exp(-x)} - 1 . \tag{A.2}$$

The function g (with range $[-2, 2]$) used in the experiments is

$$g(x) = \frac{4}{1 + \exp(-x)} - 2 . \tag{A.3}$$

A.1.1 Forward Pass.   The net input and the activation of hidden unit i are

$$net_i(t) = \sum_u w_{iu} y^u(t - 1) \tag{A.4}$$

$$y^i(t) = f_i(net_i(t)) .$$

The net input and the activation of in_j are

$$net_{in_j}(t) = \sum_u w_{in_j u} y^u(t - 1) \tag{A.5}$$

$$y^{in_j}(t) = f_{in_j}(net_{in_j}(t)) .$$

The net input and the activation of out_j are

$$net_{out_j}(t) = \sum_u w_{out_j u} y^u(t - 1) \tag{A.6}$$

$$y^{out_j}(t) = f_{out_j}(net_{out_j}(t)) .$$

The net input $net_{c_j^v}$, the internal state $s_{c_j^v}$, and the output activation $y^{c_j^v}$ of the vth memory cell of memory cell block c_j are:

$$net_{c_j^v}(t) = \sum_u w_{c_j^v u} y^u(t - 1) \tag{A.7}$$

$$s_{c_j^v}(t) = s_{c_j^v}(t - 1) + y^{in_j}(t) g\left(net_{c_j^v}(t)\right)$$

$$y^{c_j^v}(t) = y^{out_j}(t) h(s_{c_j^v}(t)) .$$

The net input and the activation of output unit k are

$$net_k(t) = \sum_{u: \ u \text{ not a gate}} w_{ku} y^u(t - 1)$$

$$y^k(t) = f_k(net_k(t)) .$$

The backward pass to be described later is based on the following truncated backpropagation formulas.

A.1.2 Approximate Derivatives for Truncated Backpropagation. The truncated version (see section 4) only approximates the partial derivatives, which is reflected by the $\approx_{tr}$ signs in the notation below. It truncates error flow once it leaves memory cells or gate units. Truncation ensures that there are no loops across which an error that left some memory cell through its input or input gate can reenter the cell through its output or output gate. This in turn ensures constant error flow through the memory cell's CEC.

In the truncated backpropagation version, the following derivatives are replaced by zero:

$$\frac{\partial net_{in_j}(t)}{\partial y^u(t-1)} \approx_{tr} 0 \ \ \forall u,$$

$$\frac{\partial net_{out_j}(t)}{\partial y^u(t-1)} \approx_{tr} 0 \ \ \forall u,$$

and

$$\frac{\partial net_{c_j}(t)}{\partial y^u(t-1)} \approx_{tr} 0 \ \ \forall u.$$

Therefore we get

$$\frac{\partial y^{in_j}(t)}{\partial y^u(t-1)} = f'_{in_j}(net_{in_j}(t))\frac{\partial net_{in_j}(t)}{\partial y^u(t-1)} \approx_{tr} 0 \ \ \forall u,$$

$$\frac{\partial y^{out_j}(t)}{\partial y^u(t-1)} = f'_{out_j}(net_{out_j}(t))\frac{\partial net_{out_j}(t)}{\partial y^u(t-1)} \approx_{tr} 0 \ \ \forall u,$$

and

$$\frac{\partial y^{c_j}(t)}{\partial y^u(t-1)} = \frac{\partial y^{c_j}(t)}{\partial net_{out_j}(t)}\frac{\partial net_{out_j}(t)}{\partial y^u(t-1)} + \frac{\partial y^{c_j}(t)}{\partial net_{in_j}(t)}\frac{\partial net_{in_j}(t)}{\partial y^u(t-1)}$$
$$+ \frac{\partial y^{c_j}(t)}{\partial net_{c_j}(t)}\frac{\partial net_{c_j}(t)}{\partial y^u(t-1)} \approx_{tr} 0 \ \ \forall u.$$

This implies for all w_{lm} not on connections to c_j^v, in_j, out_j (that is, $l \notin \{c_j^v, in_j, out_j\}$):

$$\frac{\partial y^{c_j^v}(t)}{\partial w_{lm}} = \sum_u \frac{\partial y^{c_j^v}(t)}{\partial y^u(t-1)}\frac{\partial y^u(t-1)}{\partial w_{lm}} \approx_{tr} 0.$$

The truncated derivatives of output unit k are:

$$\frac{\partial y^k(t)}{\partial w_{lm}} = f'_k(net_k(t))\left(\sum_{u:\ u \text{ not a gate}} w_{ku}\frac{\partial y^u(t-1)}{\partial w_{lm}} + \delta_{kl}y^m(t-1)\right)$$

$$
\approx_{tr} f_k'(net_k(t)) \left(\sum_j \sum_{v=1}^{S_j} \delta_{c_j^v l} w_{kc_j^v} \frac{\partial y_j^{c_j^v}(t-1)}{\partial w_{lm}} \right.
$$

$$
+ \sum_j \left(\delta_{in_j l} + \delta_{out_j l} \right) \sum_{v=1}^{S_j} w_{kc_j^v} \frac{\partial y_j^{c_j^v}(t-1)}{\partial w_{lm}}
$$

$$
\left. + \sum_{i:\ i \text{ hidden unit}} w_{ki} \frac{\partial y^i(t-1)}{\partial w_{lm}} + \delta_{kl} y^m(t-1) \right)
$$

$$
= f_k'(net_k(t)) \begin{cases} y^m(t-1) & l = k \\ w_{kc_j^v} \frac{\partial y_j^{c_j^v}(t-1)}{\partial w_{lm}} & l = c_j^v \\ \sum_{v=1}^{S_j} w_{kc_j^v} \frac{\partial y_j^{c_j^v}(t-1)}{\partial w_{lm}} & l = in_j \text{ OR } l = out_j \\ \sum_{i:\ i \text{ hidden unit}} w_{ki} \frac{\partial y^i(t-1)}{\partial w_{lm}} & l \text{ otherwise} \end{cases}
$$

$$
(A.8)
$$

where δ is the Kronecker delta ($\delta_{ab} = 1$ if $a = b$ and 0 otherwise), and S_j is the size of memory cell block c_j. The truncated derivatives of a hidden unit i that is not part of a memory cell are:

$$
\frac{\partial y^i(t)}{\partial w_{lm}} = f_i'(net_i(t)) \frac{\partial net_i(t)}{\partial w_{lm}} \approx_{tr} \delta_{li} f_i'(net_i(t)) y^m(t-1) . \tag{A.9}
$$

(Here it would be possible to use the full gradient without affecting constant error flow through internal states of memory cells.)

Cell block c_j's truncated derivatives are:

$$
\frac{\partial y^{in_j}(t)}{\partial w_{lm}} = f_{in_j}'(net_{in_j}(t)) \frac{\partial net_{in_j}(t)}{\partial w_{lm}}
$$
$$
\approx_{tr} \delta_{in_j l} f_{in_j}'(net_{in_j}(t)) y^m(t-1) . \tag{A.10}
$$

$$
\frac{\partial y^{out_j}(t)}{\partial w_{lm}} = f_{out_j}'(net_{out_j}(t)) \frac{\partial net_{out_j}(t)}{\partial w_{lm}}
$$
$$
\approx_{tr} \delta_{out_j l} f_{out_j}'(net_{out_j}(t)) y^m(t-1) . \tag{A.11}
$$

$$
\frac{\partial s_{c_j^v}(t)}{\partial w_{lm}} = \frac{\partial s_{c_j^v}(t-1)}{\partial w_{lm}}
$$
$$
+ \frac{\partial y^{in_j}(t)}{\partial w_{lm}} g\left(net_{c_j^v}(t)\right) + y^{in_j}(t) g'\left(net_{c_j^v}(t)\right) \frac{\partial net_{c_j^v}(t)}{\partial w_{lm}}
$$
$$
\approx_{tr} \left(\delta_{in_j l} + \delta_{c_j^v l} \right) \frac{\partial s_{c_j^v}(t-1)}{\partial w_{lm}} + \delta_{in_j l} \frac{\partial y^{in_j}(t)}{\partial w_{lm}} g\left(net_{c_j^v}(t)\right)
$$
$$
+ \delta_{c_j^v l} y^{in_j}(t) g'\left(net_{c_j^v}(t)\right) \frac{\partial net_{c_j^v}(t)}{\partial w_{lm}}
$$

$$
\begin{aligned}
&= \left(\delta_{in_j l} + \delta_{c_j^v l}\right) \frac{\partial s_{c_j^v}(t-1)}{\partial w_{lm}} \\
&\quad + \delta_{in_j l} f'_{in_j}(net_{in_j}(t)) \, g\left(net_{c_j^v}(t)\right) y^m(t-1) \\
&\quad + \delta_{c_j^v l} \, y^{in_j}(t) \, g'\left(net_{c_j^v}(t)\right) y^m(t-1) \, .
\end{aligned}
\tag{A.12}
$$

$$
\begin{aligned}
\frac{\partial y^{c_j^v}(t)}{\partial w_{lm}} &= \frac{\partial y^{out_j}(t)}{\partial w_{lm}} h(s_{c_j^v}(t)) + h'(s_{c_j^v}(t)) \frac{\partial s_{c_j^v}(t)}{\partial w_{lm}} y^{out_j}(t) \\
&\approx_{tr} \delta_{out_j l} \frac{\partial y^{out_j}(t)}{\partial w_{lm}} h(s_{c_j^v}(t)) \\
&\quad + \left(\delta_{in_j l} + \delta_{c_j^v l}\right) h'(s_{c_j^v}(t)) \frac{\partial s_{c_j^v}(t)}{\partial w_{lm}} y^{out_j}(t) \, .
\end{aligned}
\tag{A.13}
$$

To update the system efficiently at time t, the only (truncated) derivatives that need to be stored at time $t-1$ are

$$
\frac{\partial s_{c_j^v}(t-1)}{\partial w_{lm}},
$$

where $l = c_j^v$ or $l = in_j$.

A.1.3 Backward Pass. We will describe the backward pass only for the particularly efficient truncated gradient version of the LSTM algorithm. For simplicity we will use equal signs even where approximations are made according to the truncated backpropagation equations above.

The squared error at time t is given by

$$
E(t) = \sum_{k:\, k \text{ output unit}} \left(t^k(t) - y^k(t)\right)^2,
\tag{A.14}
$$

where $t^k(t)$ is output unit k's target at time t.

Time t's contribution to w_{lm}'s gradient-based update with learning rate α is

$$
\triangle w_{lm}(t) = -\alpha \, \frac{\partial E(t)}{\partial w_{lm}} \, .
\tag{A.15}
$$

We define some unit l's error at time step t by

$$
e_l(t) := -\frac{\partial E(t)}{\partial net_l(t)} \, .
\tag{A.16}
$$

Using (almost) standard backpropagation, we first compute updates for weights to output units ($l = k$), weights to hidden units ($l = i$) and weights

to output gates ($l = out_j$). We obtain (compare formulas A.8, A.9, and A.11):

$$l = k \text{ (output)} : \quad e_k(t) = f_k'(net_k(t)) \left(t^k(t) - y^k(t) \right) , \qquad (A.17)$$

$$l = i \text{ (hidden)} : \quad e_i(t) = f_i'(net_i(t)) \sum_{k:\ k \text{ output unit}} w_{ki} e_k(t) , \qquad (A.18)$$

$l = out_j$ (output gates) :

$$e_{out_j}(t) = f_{out_j}'(net_{out_j}(t)) \left(\sum_{v=1}^{S_j} h(s_{c_j^v}(t)) \sum_{k:\ k \text{ output unit}} w_{kc_j^v} e_k(t) \right) . \quad (A.19)$$

For all possible l time t's contribution to w_{lm}'s update is

$$\Delta w_{lm}(t) = \alpha \, e_l(t) \, y^m(t-1) . \qquad (A.20)$$

The remaining updates for weights to input gates ($l = in_j$) and to cell units ($l = c_j^v$) are less conventional. We define some internal state $s_{c_j^v}$'s error:

$$
\begin{aligned}
e_{s_{c_j^v}} &:= -\frac{\partial E(t)}{\partial s_{c_j^v}(t)} \\
&= f_{out_j}(net_{out_j}(t)) \, h'(s_{c_j^v}(t)) \sum_{k:\ k \text{ output unit}} w_{kc_j^v} e_k(t) .
\end{aligned}
\qquad (A.21)
$$

We obtain for $l = in_j$ or $l = c_j^v$, $v = 1, \dots, S_j$

$$-\frac{\partial E(t)}{\partial w_{lm}} = \sum_{v=1}^{S_j} e_{s_{c_j^v}}(t) \frac{\partial s_{c_j^v}(t)}{\partial w_{lm}} . \qquad (A.22)$$

The derivatives of the internal states with respect to weights and the corresponding weight updates are as follows (compare expression A.12):

$l = in_j$ (input gates) :
$$\frac{\partial s_{c_j^v}(t)}{\partial w_{in_j m}} = \frac{\partial s_{c_j^v}(t-1)}{\partial w_{in_j m}} + g(net_{c_j^v}(t)) f_{in_j}'(net_{in_j}(t)) y^m(t-1) ; \qquad (A.23)$$

therefore, time t's contribution to $w_{in_j m}$'s update is (compare expression A.8):

$$\Delta w_{in_j m}(t) = \alpha \sum_{v=1}^{S_j} e_{s_{c_j^v}}(t) \frac{\partial s_{c_j^v}(t)}{\partial w_{in_j m}} . \qquad (A.24)$$

Similarly we get (compare expression A.12):

$$l = c_j^v \text{ (memory cells)} :$$
$$\frac{\partial s_{c_j^v}(t)}{\partial w_{c_j^v m}} = \frac{\partial s_{c_j^v}(t-1)}{\partial w_{c_j^v m}} + g'(net_{c_j^v}(t)) f_{in_j}(net_{in_j}(t)) y^m(t-1) ; \qquad \text{(A.25)}$$

therefore time t's contribution to $w_{c_j^v m}$'s update is (compare expression A.8):

$$\Delta w_{c_j^v m}(t) = \alpha e_{s_{c_j^v}}(t) \frac{\partial s_{c_j^v}(t)}{\partial w_{c_j^v m}} . \qquad \text{(A.26)}$$

All we need to implement for the backward pass are equations A.17 through A.21 and A.23 through A.26. Each weight's total update is the sum of the contributions of all time steps.

A.1.4 Computational Complexity. LSTM's update complexity per time step is

$$O(KH + KCS + HI + CSI) = O(W), \qquad \text{(A.27)}$$

where K is the number of output units, C is the number of memory cell blocks, $S > 0$ is the size of the memory cell blocks, H is the number of hidden units, I is the (maximal) number of units forward connected to memory cells, gate units and hidden units, and

$$W = KH + KCS + CSI + 2CI + HI = O(KH + KCS + CSI + HI)$$

is the number of weights. Expression A.27 is obtained by considering all computations of the backward pass: equation A.17 needs K steps; A.18 needs KH steps; A.19 needs KSC steps; A.20 needs $K(H+C)$ steps for output units, HI steps for hidden units, CI steps for output gates; A.21 needs KCS steps; A.23 needs CSI steps; A.24 needs CSI steps; A.25 needs CSI steps; A.26 needs CSI steps. The total is $K + 2KH + KC + 2KSC + HI + CI + 4CSI$ steps, or $O(KH + KSC + HI + CSI)$ steps. We conclude that LSTM algorithm's update complexity per time step is just like BPTT's for a fully recurrent net.

At a given time step, only the $2CSI$ most recent $\partial s_{c_j^v}/\partial w_{lm}$ values from equations A.23 and A.25 need to be stored. Hence LSTM's storage complexity also is $O(W)$; it does not depend on the input sequence length.

A.2 Error Flow. We compute how much an error signal is scaled while flowing back through a memory cell for q time steps. As a by-product, this analysis reconfirms that the error flow within a memory cell's CEC is indeed constant, provided that truncated backpropagation cuts off error flow trying to leave memory cells (see also section 3.2). The analysis also highlights a

potential for undesirable long-term drifts of s_{c_j}, as well as the beneficial, countermanding influence of negatively biased input gates.

Using the truncated backpropagation learning rule, we obtain

$$
\begin{aligned}
\frac{\partial s_{c_j}(t-k)}{\partial s_{c_j}(t-k-1)} &= 1 + \frac{\partial y^{in_j}(t-k)}{\partial s_{c_j}(t-k-1)} g\left(net_{c_j}(t-k)\right) \\
&\quad + y^{in_j}(t-k)g'\left(net_{c_j}(t-k)\right) \frac{\partial net_{c_j}(t-k)}{\partial s_{c_j}(t-k-1)} \\
&= 1 + \sum_{u}\left[\frac{\partial y^{in_j}(t-k)}{\partial y^u(t-k-1)} \frac{\partial y^u(t-k-1)}{\partial s_{c_j}(t-k-1)} \right] \\
&\quad \times g\left(net_{c_j}(t-k)\right) \\
&\quad + y^{in_j}(t-k)g'\left(net_{c_j}(t-k)\right) \\
&\quad \times \sum_{u}\left[\frac{\partial net_{c_j}(t-k)}{\partial y^u(t-k-1)} \frac{\partial y^u(t-k-1)}{\partial s_{c_j}(t-k-1)} \right] \\
&\approx_{tr} 1.
\end{aligned}
\tag{A.28}
$$

The $\approx_{tr}$ sign indicates equality due to the fact that truncated backpropagation replaces by zero the following derivatives:

$$
\frac{\partial y^{in_j}(t-k)}{\partial y^u(t-k-1)} \;\; \forall u \quad \text{and} \quad \frac{\partial net_{c_j}(t-k)}{\partial y^u(t-k-1)} \;\; \forall u.
$$

In what follows, an error $\vartheta_j(t)$ starts flowing back at c_j's output. We redefine

$$
\vartheta_j(t) := \sum_{i} w_{ic_j}\vartheta_i(t+1) \,.
\tag{A.29}
$$

Following the definitions and conventions of section 3.1, we compute error flow for the truncated backpropagation learning rule. The error occurring at the output gate is

$$
\vartheta_{out_j}(t) \approx_{tr} \frac{\partial y^{out_j}(t)}{\partial net_{out_j}(t)} \frac{\partial y^{c_j}(t)}{\partial y^{out_j}(t)} \vartheta_j(t) \,.
\tag{A.30}
$$

The error occurring at the internal state is

$$
\vartheta_{s_{c_j}}(t) = \frac{\partial s_{c_j}(t+1)}{\partial s_{c_j}(t)} \vartheta_{s_{c_j}}(t+1) + \frac{\partial y^{c_j}(t)}{\partial s_{c_j}(t)} \vartheta_j(t) \,.
\tag{A.31}
$$

Since we use truncated backpropagation we have

$$
\vartheta_j(t) = \sum_{i:i \text{ no gate and no memory cell}} w_{ic_j}\vartheta_i(t+1);
$$

therefore we get

$$\frac{\partial \vartheta_j(t)}{\partial \vartheta_{s_{c_j}}(t+1)} = \sum_i w_{ic_j} \frac{\partial \vartheta_i(t+1)}{\partial \vartheta_{s_{c_j}}(t+1)} \approx_{tr} 0 \; . \tag{A.32}$$

Equations A.31 and A.32 imply constant error flow through internal states of memory cells:

$$\frac{\partial \vartheta_{s_{c_j}}(t)}{\partial \vartheta_{s_{c_j}}(t+1)} = \frac{\partial s_{c_j}(t+1)}{\partial s_{c_j}(t)} \approx_{tr} 1 \; . \tag{A.33}$$

The error occurring at the memory cell input is

$$\vartheta_{c_j}(t) = \frac{\partial g(net_{c_j}(t))}{\partial net_{c_j}(t)} \frac{\partial s_{c_j}(t)}{\partial g(net_{c_j}(t))} \vartheta_{s_{c_j}}(t) \; . \tag{A.34}$$

The error occurring at the input gate is

$$\vartheta_{in_j}(t) \approx_{tr} \frac{\partial y^{in_j}(t)}{\partial net_{in_j}(t)} \frac{\partial s_{c_j}(t)}{\partial y^{in_j}(t)} \vartheta_{s_{c_j}}(t) \; . \tag{A.35}$$

A.2.1 No External Error Flow. Errors are propagated back from units l to unit v along outgoing connections with weights w_{lv}. This "external error" (note that for conventional units there is nothing but external error) at time t is

$$\vartheta_v^e(t) = \frac{\partial y^v(t)}{\partial net_v(t)} \sum_l \frac{\partial net_l(t+1)}{\partial y^v(t)} \vartheta_l(t+1) \; . \tag{A.36}$$

We obtain

$$\begin{aligned}
\frac{\partial \vartheta_v^e(t-1)}{\partial \vartheta_j(t)} &= \frac{\partial y^v(t-1)}{\partial net_v(t-1)} \left(\frac{\partial \vartheta_{out_j}(t)}{\partial \vartheta_j(t)} \frac{\partial net_{out_j}(t)}{\partial y^v(t-1)} \right. \\
&\quad + \left. \frac{\partial \vartheta_{in_j}(t)}{\partial \vartheta_j(t)} \frac{\partial net_{in_j}(t)}{\partial y^v(t-1)} + \frac{\partial \vartheta_{c_j}(t)}{\partial \vartheta_j(t)} \frac{\partial net_{c_j}(t)}{\partial y^v(t-1)} \right) \\
&\approx_{tr} 0 \; . \tag{A.37}
\end{aligned}$$

We observe that the error ϑ_j arriving at the memory cell output is not back-propagated to units v by external connections to in_j, out_j, c_j.

A.2.2 Error Flow Within Memory Cells. We now focus on the error back-flow within a memory cell's CEC. This is actually the only type of error flow that can bridge several time steps. Suppose error $\vartheta_j(t)$ arrives at c_j's output

at time t and is propagated back for q steps until it reaches in_j or the memory cell input $g(net_{c_j})$. It is scaled by a factor of

$$\frac{\partial \vartheta_v(t-q)}{\partial \vartheta_j(t)},$$

where $v = in_j$, c_j. We first compute

$$\frac{\partial \vartheta_{s_{c_j}}(t-q)}{\partial \vartheta_j(t)} \approx_{tr} \begin{cases} \frac{\partial y^{c_j}(t)}{\partial s_{c_j}(t)} & q = 0 \\ \frac{\partial s_{c_j}(t-q+1)}{\partial s_{c_j}(t-q)} \frac{\partial \vartheta_{s_{c_j}}(t-q+1)}{\partial \vartheta_j(t)} & q > 0 \end{cases}. \qquad (A.38)$$

Expanding equation A.38, we obtain

$$\frac{\partial \vartheta_v(t-q)}{\partial \vartheta_j(t)} \approx_{tr} \frac{\partial \vartheta_v(t-q)}{\partial \vartheta_{s_{c_j}}(t-q)} \frac{\partial \vartheta_{s_{c_j}}(t-q)}{\partial \vartheta_j(t)}$$

$$\approx_{tr} \frac{\partial \vartheta_v(t-q)}{\partial \vartheta_{s_{c_j}}(t-q)} \left(\prod_{m=q}^{1} \frac{\partial s_{c_j}(t-m+1)}{\partial s_{c_j}(t-m)} \right) \frac{\partial y^{c_j}(t)}{\partial s_{c_j}(t)}$$

$$\approx_{tr} y^{out_j}(t) h'(s_{c_j}(t)) \begin{cases} g'(net_{c_j}(t-q) y^{in_j}(t-q) & v = c_j \\ g(net_{c_j}(t-q) f'_{in_j}(net_{in_j}(t-q)) & v = in_j \end{cases} \qquad (A.39)$$

Consider the factors in the previous equation's last expression. Obviously, error flow is scaled only at times t (when it enters the cell) and $t-q$ (when it leaves the cell), but not in between (constant error flow through the CEC). We observe:

1. The output gate's effect is $y^{out_j}(t)$ scales down those errors that can be reduced early during training without using the memory cell. It also scales down those errors resulting from using (activating/deactivating) the memory cell at later training stages. Without the output gate, the memory cell might, for instance, suddenly start causing avoidable errors in situations that already seemed under control (because it was easy to reduce the corresponding errors without memory cells). See "Output Weight Conflict" in section 3 and "Abuse Problem and Solution" (section 4.7).

2. If there are large positive or negative $s_{c_j}(t)$ values (because s_{c_j} has drifted since time step $t-q$), then $h'(s_{c_j}(t))$ may be small (assuming that h is a logistic sigmoid). See section 4. Drifts of the memory cell's internal state s_{c_j} can be countermanded by negatively biasing the input gate in_j (see section 4 and the next point). Recall from section 4 that the precise bias value does not matter much.

3. $y^{in_j}(t-q)$ and $f'_{in_j}(net_{in_j}(t-q))$ are small if the input gate is negatively biased (assume f_{in_j} is a logistic sigmoid). However, the potential sig-

nificance of this is negligible compared to the potential significance of drifts of the internal state s_{c_j}.

Some of the factors above may scale down LSTM's overall error flow, but not in a manner that depends on the length of the time lag. The flow will still be much more effective than an exponentially (of order q) decaying flow without memory cells.

Acknowledgments

References

Almeida, L. B. (1987). A learning rule for asynchronous perceptrons with feedback in a combinatorial environment. In *IEEE 1st International Conference on Neural Networks, San Diego* (Vol. 2, pp. 609–618).

Baldi, P., & Pineda, F. (1991). Contrastive learning and neural oscillator. *Neural Computation, 3*, 526–545.

Bengio, Y., & Frasconi, P. (1994). Credit assignment through time: Alternatives to backpropagation. In J. D. Cowan, G. Tesauro, & J. Alspector (Eds.), *Advances in neural information processing systems 6* (pp. 75–82). San Mateo, CA: Morgan Kaufmann.

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks, 5*(2), 157–166.

Cleeremans, A., Servan-Schreiber, D., & McClelland, J. L. (1989). Finite-state automata and simple recurrent networks. *Neural Computation, 1*, 372–381.

de Vries, B., & Principe, J. C. (1991). A theory for neural networks with time delays. In R. P. Lippmann, J. E. Moody, & D. S. Touretzky (Eds.), *Advances in neural information processing systems 3*, (pp. 162–168). San Mateo, CA: Morgan Kaufmann.

Doya, K. (1992). Bifurcations in the learning of recurrent neural networks. In *Proceedings of 1992 IEEE International Symposium on Circuits and Systems* (pp. 2777–2780).

Doya, K., & Yoshizawa, S. (1989). Adaptive neural oscillator using continuous-time backpropagation learning. *Neural Networks, 2*, 375–385.

Elman, J. L. (1988). *Finding structure in time* (Tech. Rep. No. CRL 8801). San Diego: Center for Research in Language, University of California, San Diego.

Fahlman, S. E. (1991). The recurrent cascade-correlation learning algorithm. In R. P. Lippmann, J. E. Moody, & D. S. Touretzky (Eds.), *Advances in neural information processing systems 3* (pp. 190–196). San Mateo, CA: Morgan Kaufmann.

Hochreiter, J. (1991). *Untersuchungen zu dynamischen neuronalen Netzen*. Diploma thesis, Institut für Informatik, Lehrstuhl Prof. Brauer, Technische Universität München. See http://www7.informatik.tu-muenchen.de/˜hochreit.

Hochreiter, S., & Schmidhuber, J. (1995). *Long short-term memory* (Tech. Rep. No. FKI-207-95). Fakultät für Informatik, Technische Universität München.

Hochreiter, S., & Schmidhuber, J. (1996). Bridging long time lags by weight guessing and "long short-term memory." In F. L. Silva, J. C. Principe, & L. B. Almeida (Eds.), *Spatiotemporal models in biological and artificial systems* (pp. 65–72). Amsterdam: IOS Press.

Hochreiter, S., & Schmidhuber, J. (1997). LSTM can solve hard long time lag problems. In *Advances in neural information processing systems 9*. Cambridge, MA: MIT Press.

Lang, K., Waibel, A., & Hinton, G. E. (1990). A time-delay neural network architecture for isolated word recognition. *Neural Networks, 3*, 23–43.

Lin, T., Horne, B. G., Tino, P., & Giles, C. L. (1996). Learning long-term dependencies in NARX recurrent neural networks. *IEEE Transactions on Neural Networks, 7*, 1329–1338.

Miller, C. B., & Giles, C. L. (1993). Experimental comparison of the effect of order in recurrent neural networks. *International Journal of Pattern Recognition and Artificial Intelligence, 7*(4), 849–872.

Mozer, M. C. (1989). A focused back-propagation algorithm for temporal sequence recognition. *Complex Systems, 3*, 349–381.

Mozer, M. C. (1992). Induction of multiscale temporal structure. In J. E. Moody, S. J. Hanson, & R. P. Lippman (Eds.), *Advances in neural information processing systems 4* (pp. 275–282). San Mateo, CA: Morgan Kaufmann.

Pearlmutter, B. A. (1989). Learning state space trajectories in recurrent neural networks. *Neural Computation, 1*(2), 263–269.

Pearlmutter, B. A. (1995). Gradient calculations for dynamic recurrent neural networks: A survey. *IEEE Transactions on Neural Networks, 6*(5), 1212–1228.

Pineda, F. J. (1987). Generalization of back-propagation to recurrent neural networks. *Physical Review Letters, 19*(59), 2229–2232.

Pineda, F. J. (1988). Dynamics and architecture for neural computation. *Journal of Complexity, 4*, 216–245.

Plate, T. A. (1993). Holographic recurrent networks. In S. J. Hanson, J. D. Cowan, & C. L. Giles (Eds.), *Advances in neural information processing systems 5* (pp. 34–41). San Mateo, CA: Morgan Kaufmann.

Pollack, J. B. (1991). Language induction by phase transition in dynamical recognizers. In R. P. Lippmann, J. E. Moody, & D. S. Touretzky (Eds.), *Advances in neural information processing systems 3* (pp. 619–626). San Mateo, CA: Morgan Kaufmann.

Puskorius, G. V., and Feldkamp, L. A. (1994). Neurocontrol of nonlinear dynamical systems with Kalman filter trained recurrent networks. *IEEE Transactions on Neural Networks, 5*(2), 279–297.

Ring, M. B. (1993). Learning sequential tasks by incrementally adding higher orders. In S. J. Hanson, J. D. Cowan, & C. L. Giles (Eds.), *Advances in neural information processing systems 5* (pp. 115–122). San Mateo, CA: Morgan Kaufmann.

Robinson, A. J., & Fallside, F. (1987). *The utility driven dynamic error propagation network* (Tech. Rep. No. CUED/F-INFENG/TR.1). Cambridge: Cambridge University Engineering Department.

Schmidhuber, J. (1989). A local learning algorithm for dynamic feedforward and recurrent networks. *Connection Science, 1*(4), 403–412.

Schmidhuber, J. (1992a). A fixed size storage $O(n^3)$ time complexity learning algorithm for fully recurrent continually running networks. *Neural Computation, 4*(2), 243–248.

Schmidhuber, J. (1992b). Learning complex, extended sequences using the principle of history compression. *Neural Computation, 4*(2), 234–242.

Schmidhuber, J. (1992c). Learning unambiguous reduced sequence descriptions. In J. E. Moody, S. J. Hanson, & R. P. Lippman (Eds.), *Advances in neural information processing systems 4* (pp. 291–298). San Mateo, CA: Morgan Kaufmann.

Schmidhuber, J. (1993). *Netzwerkarchitekturen, Zielfunktionen und Kettenregel.* Habilitationsschrift, Institut für Informatik, Technische Universität München.

Schmidhuber, J., & Hochreiter, S. (1996). *Guessing can outperform many long time lag algorithms* (Tech. Rep. No. IDSIA-19-96). Lugano, Switzerland: Instituto Dalle Molle di Studi sull'Intelligenza Artificiale.

Silva, G. X., Amaral, J. D., Langlois, T., & Almeida, L. B. (1996). Faster training of recurrent networks. In F. L. Silva, J. C. Principe, & L. B. Almeida (Eds.), *Spatiotemporal models in biological and artificial systems* (pp. 168–175). Amsterdam: IOS Press.

Smith, A. W., & Zipser, D. (1989). Learning sequential structures with the real-time recurrent learning algorithm. *International Journal of Neural Systems, 1*(2), 125–131.

Sun, G., Chen, H., & Lee, Y. (1993). Time warping invariant neural networks. In S. J. Hanson, J. D. Cowan, & C. L. Giles (Eds.), *Advances in neural information processing systems 5* (pp. 180–187). San Mateo, CA: Morgan Kaufmann.

Watrous, R. L., & Kuhn, G. M. (1992). Induction of finite-state languages using second-order recurrent networks. *Neural Computation, 4*, 406–414.

Werbos, P. J. (1988). Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks, 1*, 339–356.

Williams, R. J. (1989). *Complexity of exact gradient computation algorithms for recurrent neural networks* (Tech. Rep. No. NU-CCS-89-27). Boston: Northeastern University, College of Computer Science.

Williams, R. J. & Peng, J. (1990). An efficient gradient-based algorithm for on-line training of recurrent network trajectories. *Neural Computation, 4*, 491–501.

Williams, R. J., & Zipser, D. (1992). Gradient-based learning algorithms for recurrent networks and their computational complexity. In Y. Chauvin, & D. E. Rumelhart (Eds.), *Back-propagation: Theory, architectures and applications.* Hillsdale, NJ: Erlbaum.

This article has been cited by:

1. Qingchen Zhang, Laurence T. Yang, Zhikui Chen, Peng Li. 2018. A survey on deep learning for big data. *Information Fusion* **42**, 146-157. [Crossref]

2. Pil-Soo Kim, Dong-Gyu Lee, Seong-Whan Lee. 2018. Discriminative context learning with gated recurrent unit for group activity recognition. *Pattern Recognition* **76**, 149-161. [Crossref]

3. Chengcheng Wan, Yanmin Zhu, Jiadi Yu, Yanyan Shen. 2018. SMOPAT: Mining semantic mobility patterns from trajectories of private vehicles. *Information Sciences* **429**, 12-25. [Crossref]

4. Rui Zhang, Zhigang Jin, Xiaohui Liu. 2018. A Study on the Analysis Model of the Ranking of the Theme of Weibo. *International Journal of Pattern Recognition and Artificial Intelligence* **32**:03, 1851003. [Crossref]

5. Nastaran Mohammadian Rad, Seyed Mostafa Kia, Calogero Zarbo, Twan van Laarhoven, Giuseppe Jurman, Paola Venuti, Elena Marchiori, Cesare Furlanello. 2018. Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders. *Signal Processing* **144**, 180-191. [Crossref]

6. Ruibo Yan, Xi Xiao, Guangwu Hu, Sancheng Peng, Yong Jiang. 2018. New deep learning method to detect code injection attacks on hybrid applications. *Journal of Systems and Software* **137**, 67-77. [Crossref]

7. Qiqi Hou, Jinjun Wang, Ruibin Bai, Sanping Zhou, Yihong Gong. 2018. Face alignment recurrent network. *Pattern Recognition* **74**, 448-458. [Crossref]

8. Xinggang Wang, Yongluan Yan, Peng Tang, Xiang Bai, Wenyu Liu. 2018. Revisiting multiple instance neural networks. *Pattern Recognition* **74**, 15-24. [Crossref]

9. Shiluo Xu, Ruiqing Niu. 2018. Displacement prediction of Baijiabao landslide based on empirical mode decomposition and long short-term memory neural network in Three Gorges area, China. *Computers & Geosciences* **111**, 87-96. [Crossref]

10. Zhang-Ling Chen, Jun Wang, Wen-Juan Li, Nan Li, Hua-Ming Wu, Da-Wei Wang. 2018. Convolutional neural network with nonlinear competitive units. *Signal Processing: Image Communication* **60**, 193-198. [Crossref]

11. Rui Santos, Patricia Murrieta-Flores, Pável Calado, Bruno Martins. 2018. Toponym matching through deep neural networks. *International Journal of Geographical Information Science* **32**:2, 324-348. [Crossref]

12. Seung Hyun Kim, Su Chang Lim, Do Yeon Kim. 2018. Intelligent intrusion detection system featuring a virtual fence, active intruder detection, classification, tracking, and action recognition. *Annals of Nuclear Energy* **112**, 845-855. [Crossref]

13. Fang Zhao, Jiashi Feng, Jian Zhao, Wenhan Yang, Shuicheng Yan. 2018. Robust LSTM-Autoencoders for Face De-Occlusion in the Wild. *IEEE Transactions on Image Processing* **27**:2, 778-790. [Crossref]

14. Brian McFee. Statistical Methods for Scene and Event Classification 103-146. [Crossref]

15. Yury S. Prostov, Yury V. Tiumentsev. **736**, 33. [Crossref]

16. Dinh Nguyen, Khuong Vo, Dang Pham, Mao Nguyen, Tho Quan. A Deep Architecture for Sentiment Analysis of News Articles 129-140. [Crossref]

17. Jan Jakubik. Evaluation of Gated Recurrent Neural Networks in Music Classification Tasks 27-37. [Crossref]

18. Prachya Boonkwan, Thepchai Supnithi. Bidirectional Deep Learning of Context Representation for Joint Word Segmentation and POS Tagging 184-196. [Crossref]

19. Marvin Coto-Jiménez, John Goddard-Close. 2018. LSTM Deep Neural Networks Postfiltering for Enhancing Synthetic Voices. *International Journal of Pattern Recognition and Artificial Intelligence* **32**:01, 1860008. [Crossref]

20. Jorge Revuelta Herrero, Álvaro Lozano Murciego, Alberto López Barriuso, Daniel Hernández de la Iglesia, Gabriel Villarrubia González, Juan Manuel Corchado Rodríguez, Rita Carreira. Non Intrusive Load Monitoring (NILM): A State of the Art 125-138. [Crossref]

21. Seyed Sajad Mousavi, Michael Schukat, Enda Howley. Deep Reinforcement Learning: An Overview 426-440. [Crossref]

22. Sang Il Lee, Seong Joon Yoo. A New Method for Portfolio Construction Using a Deep Predictive Model 260-266. [Crossref]

23. Kai Xu, Zhanfan Zhou, Tianyong Hao, Wenyin Liu. **639**, 355. [Crossref]

24. Kriti Khare, Helen Lam, Anshuman Khare. Educational Data Mining (EDM): Researching Impact on Online Business Education 37-53. [Crossref]

25. Sou Miyamoto, Takashi Nose, Suzunosuke Ito, Harunori Koike, Yuya Chiba, Akinori Ito, Takahiro Shinozaki. Voice Conversion from Arbitrary Speakers Based on Deep Neural Networks with Adversarial Learning 97-103. [Crossref]

26. Mohammed Jabreel, Fadi Hassan, Antonio Moreno. Target-Dependent Sentiment Analysis of Tweets Using Bidirectional Gated Recurrent Neural Networks 39-55. [Crossref]

27. Buyi Yin, Yingmin Jia. Visual Based Abnormal Target Annotation with Recurrent Neural Networks 155-167. [Crossref]

28. V. G. Sujadevi, K. P. Soman, R. Vinayakumar. Real-Time Detection of Atrial Fibrillation from Short Time Single Lead ECG Traces Using Recurrent Neural Networks 212-221. [Crossref]

29. Lingyu Zhu, Juha Kesseli, Matti Nykter, Heikki Huttunen. Predicting Gene Expression Levels from Histone Modification Signals with Convolutional Recurrent Neural Networks 555-558. [Crossref]

30. Felix Kunde, Alexander Hartenstein, Petra Sauer. 253. [Crossref]

31. Eva Almansa, David Anseán, Inés Couso, Luciano Sánchez. Health Assessment of Automotive Batteries Through Computational Intelligence-Based Soft Sensors: An Empirical Study 47-56. [Crossref]

32. Mladen Karan, Jan Šnajder. 2018. Paraphrase-focused learning to rank for domain-specific frequently asked questions retrieval. *Expert Systems with Applications* **91**, 418-433. [Crossref]

33. Aleksandr Bakhshiev, Lev Stankevich. **736**, 47. [Crossref]

34. Manuele Palmeri, Filippo Vella, Ignazio Infantino, Salvatore Gaglio. Sign Languages Recognition Based on Neural Network Architecture 109-118. [Crossref]

35. Humphrey Sheil, Omer Rana. Classifying and Recommending Using Gradient Boosted Machines and Vector Space Models 214-221. [Crossref]

36. Jayadeep Pati, Krishnkant Swarnkar, Gourav Dhakad, K. K. Shukla. Temporal Modelling of Bug Numbers of Open Source Software Applications Using LSTM 189-203. [Crossref]

37. Jônatas Wehrmann, Gabriel S. Simões, Rodrigo C. Barros, Victor F. Cavalcante. 2018. Adult content detection in videos with convolutional and recurrent neural networks. *Neurocomputing* **272**, 432-438. [Crossref]

38. Dazhen Lin, Lingxiao Li, Donglin Cao, Yanping Lv, Xiao Ke. 2018. Multi-modality weakly labeled sentiment learning based on Explicit Emotion Signal for Chinese microblog. *Neurocomputing* **272**, 258-269. [Crossref]

39. Thomas Huybrechts, Yon Vanommeslaeghe, Dries Blontrock, Gregory Van Barel, Peter Hellinckx. Automatic Reverse Engineering of CAN Bus Data Using Machine Learning Techniques 751-761. [Crossref]

40. Luming Li, Xinhua Jiang. Predicting the Travel Time in Using Recurrent Neural Networks: A Case Study of Fuzhou 28-35. [Crossref]

41. Julià Camps, Albert Samà, Mario Martín, Daniel Rodríguez-Martín, Carlos Pérez-López, Joan M. Moreno Arostegui, Joan Cabestany, Andreu Català, Sheila Alcaine, Berta Mestre, Anna Prats, Maria C. Crespo-Maraver, Timothy J. Counihan, Patrick Browne, Leo R. Quinlan, Gearóid Ó Laighin, Dean Sweeney, Hadas Lewy, Gabriel Vainstein, Alberto Costa, Roberta Annicchiarico, Àngels Bayés, Alejandro Rodríguez-Molinero. 2018. Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit. *Knowledge-Based Systems* **139**, 119-131. [Crossref]

42. T. Theodoridis, V. Solachidis, N. Vretos, P. Daras. Human Fall Detection from Acceleration Measurements Using a Recurrent Neural Network 145-149. [Crossref]

43. Ayman El-Kilany, Amr Azzam, Samhaa R. El-Beltagy. Using Deep Neural Networks for Extracting Sentiment Targets in Arabic Tweets 3-15. [Crossref]

44. Aleksei Pugachev, Oleg Akhtiamov, Alexey Karpov, Wolfgang Minker. Deep Learning for Acoustic Addressee Detection in Spoken Dialogue Systems 45-53. [Crossref]

45. Nikita Markovnikov, Irina Kipyatkova, Alexey Karpov, Andrey Filchenkov. Deep Neural Networks in Russian Speech Recognition 54-67. [Crossref]

46. Oxana Verkholyak, Alexey Karpov. Combined Feature Representation for Emotion Classification from Russian Speech 68-73. [Crossref]

47. The Anh Le, Mikhail Y. Arkhipov, Mikhail S. Burtsev. Application of a Hybrid Bi-LSTM-CRF Model to the Task of Russian Named Entity Recognition 91-103. [Crossref]

48. . References 801-845. [Crossref]

49. Marco Gori. Learning and Reasoning With Constraints 340-444. [Crossref]

50. . Bibliography 534-551. [Crossref]

51. Xiaoqiang Zhou, Baotian Hu, Qingcai Chen, Xiaolong Wang. 2018. Recurrent convolutional neural network for answer selection in community question answering. *Neurocomputing* **274**, 8-18. [Crossref]

52. Sumeet Trehan, Kevin T. Carlberg, Louis J. Durlofsky. 2017. Error modeling for surrogates of dynamical systems using machine learning. *International Journal for Numerical Methods in Engineering* **112**:12, 1801-1827. [Crossref]

53. Mikel L. Forcada. 2017. Making sense of neural machine translation. *Translation Spaces* **6**:2, 291-309. [Crossref]

54. Kangwook Lee, Sanggyu Han, Sung-Hyon Myaeng. 2017. A discourse-aware neural network-based text model for document-level text classification. *Journal of Information Science* **3**, 016555151774364. [Crossref]

55. Wei Zheng, Hongfei Lin, Ling Luo, Zhehuan Zhao, Zhengguang Li, Yijia Zhang, Zhihao Yang, Jian Wang. 2017. An attention-based effective neural model for drug-drug interactions extraction. *BMC Bioinformatics* **18**:1. . [Crossref]

56. A. Benou, R. Veksler, A. Friedman, T. Riklin Raviv. 2017. Ensemble of expert deep neural networks for spatio-temporal denoising of contrast-enhanced MRI sequences. *Medical Image Analysis* **42**, 145-159. [Crossref]

57. David Belo, João Rodrigues, João R. Vaz, Pedro Pezarat-Correia, Hugo Gamboa. 2017. Biosignals learning and synthesis using deep neural networks. *BioMedical Engineering OnLine* **16**:1. . [Crossref]

58. Hongyang Xue, Zhou Zhao, Deng Cai. 2017. Unifying the Video and Question Attentions for Open-Ended Video Question Answering. *IEEE Transactions on Image Processing* **26**:12, 5656-5666. [Crossref]

59. Eleni Tsironi, Pablo Barros, Cornelius Weber, Stefan Wermter. 2017. An analysis of Convolutional Long Short-Term Memory Recurrent Neural Networks for gesture recognition. *Neurocomputing* **268**, 76-86. [Crossref]

60. Timo Nachstedt, Christian Tetzlaff. 2017. Working Memory Requires a Combination of Transient and Attractor-Dominated Dynamics to Process Unreliably Timed Inputs. *Scientific Reports* **7**:1. . [Crossref]

61. Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, Clara I. Sánchez. 2017. A survey on deep learning in medical image analysis. *Medical Image Analysis* **42**, 60-88. [Crossref]

62. Lian-sheng Li, Sheng-jiang Gan, Xiang-dong Yin. 2017. Feedback recurrent neural network-based embedded vector and its application in topic model. *EURASIP Journal on Embedded Systems* **2017**:1. . [Crossref]

63. Shumin Li, Junjie Chen, Bin Liu. 2017. Protein remote homology detection based on bidirectional long short-term memory. *BMC Bioinformatics* **18**:1. . [Crossref]

64. Vataya Chunwijitra, Chai Wutiwiwatchai. 2017. Classification-based spoken text selection for LVCSR language modeling. *EURASIP Journal on Audio, Speech, and Music Processing* **2017**:1. . [Crossref]

65. Fei Li, Meishan Zhang, Guohong Fu, Donghong Ji. 2017. A neural joint model for entity and relation extraction from biomedical text. *BMC Bioinformatics* **18**:1. . [Crossref]

66. Marcus Olivecrona, Thomas Blaschke, Ola Engkvist, Hongming Chen. 2017. Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics* **9**:1. . [Crossref]

67. Zhenghua Chen, Rui Zhao, Qingchang Zhu, Mustafa K. Masood, Yeng Chai Soh, Kezhi Mao. 2017. Building Occupancy Estimation with Environmental Sensors via CDBLSTM. *IEEE Transactions on Industrial Electronics* **64**:12, 9549-9559. [Crossref]

68. Mateusz Malinowski, Marcus Rohrbach, Mario Fritz. 2017. Ask Your Neurons: A Deep Learning Approach to Visual Question Answering. *International Journal of Computer Vision* **125**:1-3, 110-135. [Crossref]

69. Chen Lyu, Bo Chen, Yafeng Ren, Donghong Ji. 2017. Long short-term memory RNN for biomedical named entity recognition. *BMC Bioinformatics* **18**:1. . [Crossref]

70. Xiang Li, Ling Peng, Xiaojing Yao, Shaolong Cui, Yuan Hu, Chengzeng You, Tianhe Chi. 2017. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental Pollution* **231**, 997-1004. [Crossref]

71. Mengwen Liu, Yi Fang, Alexander G. Choulos, Dae Hoon Park, Xiaohua Hu. 2017. Product review summarization through question retrieval and diversification. *Information Retrieval Journal* **20**:6, 575-605. [Crossref]

72. Chen Liu, Weiyan Hou, Deyin Liu. 2017. Foreign Exchange Rates Forecasting with Convolutional Neural Network. *Neural Processing Letters* **46**:3, 1095-1119. [Crossref]

73. Jintao Ke, Hongyu Zheng, Hai Yang, Xiqun (Michael) Chen. 2017. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging Technologies* **85**, 591-608. [Crossref]

74. Alexander G. Ororbia II, Tomas Mikolov, David Reitter. 2017. Learning Simpler Language Models with the Differential State Framework. *Neural Computation* **29**:12, 3327-3352. [Abstract] [Full Text] [PDF] [PDF Plus]

75. Hao Tang, Liang Lu, Lingpeng Kong, Kevin Gimpel, Karen Livescu, Chris Dyer, Noah A. Smith, Steve Renals. 2017. End-to-End Neural Segmental Models for Speech Recognition. *IEEE Journal of Selected Topics in Signal Processing* **11**:8, 1254-1264. [Crossref]

76. Kun Fu, Junqi Jin, Runpeng Cui, Fei Sha, Changshui Zhang. 2017. Aligning Where to See and What to Tell: Image Captioning with Region-Based Attention and Scene-Specific Contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**:12, 2321-2334. [Crossref]

77. Panagiotis Tzirakis, George Trigeorgis, Mihalis A. Nicolaou, Bjorn W. Schuller, Stefanos Zafeiriou. 2017. End-to-End Multimodal Emotion Recognition Using Deep Neural Networks. *IEEE Journal of Selected Topics in Signal Processing* **11**:8, 1301-1309. [Crossref]

78. Wayne Xiong, Jasha Droppo, Xuedong Huang, Frank Seide, Michael L. Seltzer, Andreas Stolcke, Dong Yu, Geoffrey Zweig. 2017. Toward Human Parity in Conversational Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:12, 2410-2423. [Crossref]

79. Shinji Watanabe, Takaaki Hori, Suyoun Kim, John R. Hershey, Tomoki Hayashi. 2017. Hybrid CTC/Attention Architecture for End-to-End Speech Recognition. *IEEE Journal of Selected Topics in Signal Processing* **11**:8, 1240-1253. [Crossref]

80. Biao Zhang, Deyi Xiong, Jinsong Su, Hong Duan. 2017. A Context-Aware Recurrent Encoder for Neural Machine Translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:12, 2424-2432. [Crossref]

81. Vivienne Sze, Yu-Hsin Chen, Tien-Ju Yang, Joel S. Emer. 2017. Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proceedings of the IEEE* **105**:12, 2295-2329. [Crossref]

82. Quanzeng You, Ran Pang, Liangliang Cao, Jiebo Luo. 2017. Image-Based Appraisal of Real Estate Properties. *IEEE Transactions on Multimedia* **19**:12, 2751-2759. [Crossref]

83. Myungjong Kim, Beiming Cao, Ted Mau, Jun Wang. 2017. Speaker-Independent Silent Speech Recognition From Flesh-Point Articulatory Movements Using an LSTM Neural Network. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:12, 2323-2336. [Crossref]

84. Jose A. Gonzalez, Lam A. Cheah, Angel M. Gomez, Phil D. Green, James M. Gilbert, Stephen R. Ell, Roger K. Moore, Ed Holdsworth. 2017. Direct Speech Reconstruction From Articulatory Sensor Data by Machine Learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:12, 2362-2374. [Crossref]

85. Patrick Doetsch, Mirko Hannemann, Ralf Schluter, Hermann Ney. 2017. Inverted Alignments for End-to-End Automatic Speech Recognition. *IEEE Journal of Selected Topics in Signal Processing* **11**:8, 1265-1273. [Crossref]

86. Wei-Hung Weng, Kavishwar B. Wagholikar, Alexa T. McCray, Peter Szolovits, Henry C. Chueh. 2017. Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach. *BMC Medical Informatics and Decision Making* **17**:1. . [Crossref]

87. Jordan Carlson, Bo Liu, James Sallis, Jacqueline Kerr, J. Hipp, Vincent Staggs, Amy Papa, Kelsey Dean, Nuno Vasconcelos. 2017. Automated Ecological Assessment of Physical Activity: Advancing Direct Observation. *International Journal of Environmental Research and Public Health* **14**:12, 1487. [Crossref]

88. Iñigo Jauregi Unanue, Ehsan Zare Borzeshi, Massimo Piccardi. 2017. Recurrent neural networks with specialized word embeddings for health-domain named-entity recognition. *Journal of Biomedical Informatics* **76**, 102-109. [Crossref]

89. Jônatas Wehrmann, Anderson Mattjie, Rodrigo C. Barros. 2017. Order embeddings and character-level convolutions for multimodal alignment. *Pattern Recognition Letters* . [Crossref]

90. Marc Bolaños, Álvaro Peris, Francisco Casacuberta, Sergi Soler, Petia Radeva. 2017. Egocentric Video Description based on Temporally-Linked Sequences. *Journal of Visual Communication and Image Representation* . [Crossref]

91. Dan R. Ghica, Khulood Alyahya. 2017. On the Learnability of Programming Language Semantics. *Electronic Proceedings in Theoretical Computer Science* **261**, 57-75. [Crossref]

92. Saptarshi Misra, Sudeshna Sarkar, Pabitra Mitra. 2017. Statistical downscaling of precipitation using long short-term memory recurrent neural networks. *Theoretical and Applied Climatology* **28**. . [Crossref]

93. Seokhoon Kang. 2017. Outgoing call recommendation using neural network. *Soft Computing* **16**. . [Crossref]

94. Dylan R. Muir. Feedforward Approximations to Dynamic Recurrent Network Architectures. *Neural Computation*, ahead of print1-22. [Abstract] [PDF] [PDF Plus]

95. Hongshen Chen, Xiaorui Liu, Dawei Yin, Jiliang Tang. 2017. A Survey on Dialogue Systems. *ACM SIGKDD Explorations Newsletter* **19**:2, 25-35. [Crossref]

96. Roelant A. Stegmann, Indrė Žliobaitė, Tuukka Tolvanen, Jaakko Hollmén, Jesse Read. 2017. A survey of evaluation methods for personal route and destination prediction from mobility traces. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **14**, e1237. [Crossref]

97. Kan Chen, Rama Kovvuri, Jiyang Gao, Ram Nevatia. 2017. MSRC: multimodal spatial regression with semantic context for phrase grounding. *International Journal of Multimedia Information Retrieval* **9**. . [Crossref]

98. Shang Gao, Michael T Young, John X Qiu, Hong-Jun Yoon, James B Christian, Paul A Fearn, Georgia D Tourassi, Arvind Ramanthan. 2017. Hierarchical attention networks for information extraction from cancer pathology reports. *Journal of the American Medical Informatics Association* . [Crossref]

99. Kuai Fang, Chaopeng Shen, Daniel Kifer, Xiao Yang. 2017. Prolongation of SMAP to Spatiotemporally Seamless Coverage of Continental U.S. Using a Deep Learning Neural Network. *Geophysical Research Letters* **44**:21, 11,030-11,039. [Crossref]

100. Mohammad Sadegh Rasooli, Noura Farra, Axinia Radeva, Tao Yu, Kathleen McKeown. 2017. Cross-lingual sentiment transfer with limited resources. *Machine Translation* **10**. . [Crossref]

101. Tapomayukh Bhattacharjee, James M. Rehg, Charles C. Kemp. 2017. Inferring Object Properties with a Tactile-Sensing Array Given Varying Joint Stiffness and Velocity. *International Journal of Humanoid Robotics* **1**, 1750024. [Crossref]

102. Yusuke Goutsu, Wataru Takano, Yoshihiko Nakamura. 2017. Classification of Multi-class Daily Human Motion using Discriminative Body Parts and Sentence Descriptions. *International Journal of Computer Vision* **17**. . [Crossref]

103. S. O. Oprea, A. Garcia-Garcia, S. Orts-Escolano, V. Villena-Martinez, J. A. Castro-Vargas. 2017. A long short-term memory based Schaeffer gesture recognition system. *Expert Systems* **7**, e12247. [Crossref]

104. Tian Zhou, Juan Pablo Wachs. 2017. Early prediction for physical human robot collaboration in the operating room. *Autonomous Robots* **32**. . [Crossref]

105. Ji Gan, Weiqiang Wang. 2017. In-air handwritten English word recognition using attention recurrent translator. *Neural Computing and Applications* **46**. . [Crossref]

106. Xiyao Cheng, Ying Chen, Bixiao Cheng, Shoushan Li, Guodong Zhou. 2017. An Emotion Cause Corpus for Chinese Microblogs with Multiple-User Structures. *ACM Transactions on Asian and Low-Resource Language Information Processing* **17**:1, 1-19. [Crossref]

107. Anvita Gupta, Alex T. Müller, Berend J. H. Huisman, Jens A. Fuchs, Petra Schneider, Gisbert Schneider. 2017. Generative Recurrent Networks for De Novo Drug Design. *Molecular Informatics* **432**. . [Crossref]

108. Young-Bum Kim, Karl Stratos, Ruhi Sarikaya. 2017. A Framework for pre-training hidden-unit conditional random fields and its extension to long short term memory networks. *Computer Speech & Language* **46**, 311-326. [Crossref]

109. Marta R. Costa-jussà, Alexandre Allauzen, Loïc Barrault, Kyunghun Cho, Holger Schwenk. 2017. Introduction to the special issue on deep learning approaches for machine translation. *Computer Speech & Language* **46**, 367-373. [Crossref]

110. Huiwen Guo, Xinyu Wu, Wei Feng. 2017. Multi-stream deep networks for human action classification with sequential tensor decomposition. *Signal Processing* **140**, 198-206. [Crossref]

111. Seppo Enarvi, Peter Smit, Sami Virpioja, Mikko Kurimo. 2017. Automatic Speech Recognition With Very Large Conversational Finnish and Estonian Vocabularies. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:11, 2085-2097. [Crossref]

112. Tomoki Hayashi, Shinji Watanabe, Tomoki Toda, Takaaki Hori, Jonathan Le Roux, Kazuya Takeda. 2017. Duration-Controlled LSTM for Polyphonic Sound Event Detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:11, 2059-2070. [Crossref]

113. Ying Zheng, Hongxun Yao, Xiaoshuai Sun, Xuesong Jiang, Fatih Porikli. 2017. Breaking video into pieces for action recognition. *Multimedia Tools and Applications* **76**:21, 22195-22212. [Crossref]

114. Degen Huang, Zhenchao Jiang, Li Zou, Lishuang Li. 2017. Drug?drug interaction extraction from biomedical literature using support vector machine and long short term memory networks. *Information Sciences* **415-416**, 100-109. [Crossref]

115. Samuel Rönnqvist, Peter Sarlin. 2017. Bank distress in the news: Describing events through deep learning. *Neurocomputing* **264**, 57-70. [Crossref]

116. Janez Starc, Dunja Mladenić. 2017. Constructing a Natural Language Inference dataset using generative neural networks. *Computer Speech & Language* **46**, 94-112. [Crossref]

117. Ya Li, Jianhua Tao, Linlin Chao, Wei Bao, Yazhu Liu. 2017. CHEAVD: a Chinese natural emotional audio–visual database. *Journal of Ambient Intelligence and Humanized Computing* **8**:6, 913-924. [Crossref]

118. Shakti P. Rath. 2017. Scalable algorithms for unsupervised clustering of acoustic data for speech recognition. *Computer Speech & Language* **46**, 233-248. [Crossref]

119. Eunwoo Song, Frank K. Soong, Hong-Goo Kang. 2017. Effective Spectral and Excitation Modeling Techniques for LSTM-RNN-Based Speech Synthesis Systems. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:11, 2152-2161. [Crossref]

120. Wenhuan Lu, Ju Zhang, Xinli Zhao, Jianrong Wang, Jianwu Dang. 2017. Multimodal sensory fusion for soccer robot self-localization based on long short-term memory recurrent neural network. *Journal of Ambient Intelligence and Humanized Computing* **8**:6, 885-893. [Crossref]

121. Zhiwen Xie, Zhao Zeng, Guangyou Zhou, Weijun Wang. 2017. Topic enhanced deep structured semantic models for knowledge base question answering. *Science China Information Sciences* **60**:11. . [Crossref]

122. Yuhan Jia, Jianping Wu, Moshe Ben-Akiva, Ravi Seshadri, Yiman Du. 2017. Rainfall-integrated traffic speed prediction using deep learning method. *IET Intelligent Transport Systems* **11**:9, 531-536. [Crossref]

123. José Juan Almagro Armenteros, Casper Kaae Sønderby, Søren Kaae Sønderby, Henrik Nielsen, Ole Winther. 2017. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33**:21, 3387-3395. [Crossref]

124. Eiji Mizutani, Stuart Dreyfus. 2017. Totally model-free actor-critic recurrent neural-network reinforcement learning in non-Markovian domains. *Annals of Operations Research* **258**:1, 107-131. [Crossref]

125. Zhiling Guo, Qi Chen, Guangming Wu, Yongwei Xu, Ryosuke Shibasaki, Xiaowei Shao. 2017. Village Building Identification Based on Ensemble Convolutional Neural Networks. *Sensors* **17**:11, 2487. [Crossref]

126. Alexandros Iosifidis, Anastasios Tefas, Ioannis Pitas, Moncef Gabbouj. 2017. Big Media Data Analysis. *Signal Processing: Image Communication* **59**, 105-108. [Crossref]

127. Xi Zhang, Jiawei Shi, Di Wang, Binxing Fang. 2017. Exploiting investors social network for stock prediction in China's market. *Journal of Computational Science* . [Crossref]

128. Wonjun Hwang, Soo-Chul Lim. 2017. Inferring Interaction Force from Visual Information without Using Physical Force Sensors. *Sensors* **17**:11, 2455. [Crossref]

129. Jaehyun Ahn, Dongil Shin, Kyuho Kim, Jihoon Yang. 2017. Indoor Air Quality Analysis Using Deep Learning with Sensor Data. *Sensors* **17**:11, 2476. [Crossref]

130. Sheng Yu, Yun Cheng, Li Xie, Zhiming Luo, Min Huang, Shaozi Li. 2017. A novel recurrent hybrid network for feature fusion in action recognition. *Journal of Visual Communication and Image Representation* **49**, 192-203. [Crossref]

131. Gabriel Chartrand, Phillip M. Cheng, Eugene Vorontsov, Michal Drozdzal, Simon Turcotte, Christopher J. Pal, Samuel Kadoury, An Tang. 2017. Deep Learning: A Primer for Radiologists. *RadioGraphics* **37**:7, 2113-2131. [Crossref]

132. Robert Malouf. 2017. Abstractive morphological learning with a recurrent neural network. *Morphology* **27**:4, 431-458. [Crossref]

133. Yaming Sun, Zhenzhou Ji, Lei Lin, Xiaolong Wang, Duyu Tang. 2017. Entity disambiguation with memory network. *Neurocomputing* . [Crossref]

134. Zheng Xiao, Xiong Li, Le Wang, Qiuwei Yang, Jiayi Du, Arun Kumar Sangaiah. 2017. Using convolution control block for Chinese sentiment analysis. *Journal of Parallel and Distributed Computing* . [Crossref]

135. Yiming Li, Baogang Wei, Yonghuai Liu, Liang Yao, Hui Chen, Jifang Yu, Wenhao Zhu. 2017. Incorporating Knowledge into neural network for text representation. *Expert Systems with Applications* . [Crossref]

136. Duo Zhang, Erlend Skullestad Hølland, Geir Lindholm, Harsha Ratnaweera. 2017. Hydraulic modeling and deep learning based flow forecasting for optimizing inter catchment wastewater transfer. *Journal of Hydrology* . [Crossref]

137. Duc Tran, Hieu Mac, Van Tong, Hai Anh Tran, Linh Giang Nguyen. 2017. A LSTM based framework for handling multiclass imbalance in DGA botnet detection. *Neurocomputing* . [Crossref]

138. Roberto Bonfigli, Andrea Felicetti, Emanuele Principi, Marco Fagiani, Stefano Squartini, Francesco Piazza. 2017. Denoising Autoencoders for Non-Intrusive Load Monitoring: Improvements and Comparative Evaluation. *Energy and Buildings* . [Crossref]

139. Stefan Oehmcke, Oliver Zielinski, Oliver Kramer. 2017. Input Quality Aware Convolutional LSTM Networks for Virtual Marine Sensors. *Neurocomputing* . [Crossref]

140. Zhipeng Jiang, Chao Zhao, Bin He, Yi Guan, Jingchi Jiang. 2017. De-identification of medical records using conditional random fields and long short-term memory networks. *Journal of Biomedical Informatics* **75**, S43-S53. [Crossref]

141. Tung Tran, Ramakanth Kavuluru. 2017. Predicting mental conditions based on "history of present illness" in psychiatric notes with deep neural networks. *Journal of Biomedical Informatics* **75**, S138-S148. [Crossref]

142. Duo Zhang, Geir Lindholm, Harsha Ratnaweera. 2017. Use Long Short-Term Memory to Enhance Internet of Things for Combined Sewer Overflow Monitoring. *Journal of Hydrology* . [Crossref]

143. Akara Supratak, Hao Dong, Chao Wu, Yike Guo. 2017. DeepSleepNet: A Model for Automatic Sleep Stage Scoring Based on Raw Single-Channel EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25**:11, 1998-2008. [Crossref]

144. Shiyang Yan, Jeremy S. Smith, Wenjin Lu, Bailing Zhang. 2017. Hierarchical Multi-scale Attention Networks for action recognition. *Signal Processing: Image Communication* . [Crossref]

145. Upinder S. Bhalla. 2017. Dendrites, deep learning, and sequences in the hippocampus. *Hippocampus* **9**. . [Crossref]

146. Connor Schenck, Dieter Fox. 2017. Perceiving and reasoning about liquids using fully convolutional networks. *The International Journal of Robotics Research* **17**, 027836491773405. [Crossref]

147. Jiageng Feng, Songyang Zhang, Jun Xiao. 2017. Explorations of skeleton features for LSTM-based action recognition. *Multimedia Tools and Applications* **5**. . [Crossref]

148. Peng Xia, Jie Hu, Yinghong Peng. 2017. EMG-Based Estimation of Limb Movement Using Deep Learning With Recurrent Convolutional Neural Networks. *Artificial Organs* **22**. . [Crossref]

149. Chunyan Xu, Jian Yang, Junbin Gao. 2017. Coupled-learning convolutional neural networks for object recognition. *Multimedia Tools and Applications* **9**. . [Crossref]

150. Yijia Zhang, Wei Zheng, Hongfei Lin, Jian Wang, Zhihao Yang, Michel Dumontier. 2017. Drug–drug interaction extraction via hierarchical RNNs on sequence and shortest dependency paths. *Bioinformatics* . [Crossref]

151. Bowen Liu, Bharath Ramsundar, Prasad Kawthekar, Jade Shi, Joseph Gomes, Quang Luu Nguyen, Stephen Ho, Jack Sloane, Paul Wender, Vijay Pande. 2017. Retrosynthetic Reaction Prediction Using Neural Sequence-to-Sequence Models. *ACS Central Science* **3**:10, 1103-1113. [Crossref]

152. Shun-ichi Amari, Tomoko Ozeki, Ryo Karakida, Yuki Yoshida, Masato Okada. Dynamics of Learning in MLP: Natural Gradient and Singularity Revisited. *Neural Computation*, ahead of print1-33. [Abstract] [PDF] [PDF Plus]

153. Xin Li, Liang Yang, Yi Li, Chuan Zhou. 2017. Deep trajectory: a deep learning approach for mobile advertising in vehicular networks. *Neural Computing and Applications* **10**. . [Crossref]

154. Amir Farzad, Hoda Mashayekhi, Hamid Hassanpour. 2017. A comparative performance analysis of different activation functions in LSTM networks for classification. *Neural Computing and Applications* **9**. . [Crossref]

155. Sen Wang, Ronald Clark, Hongkai Wen, Niki Trigoni. 2017. End-to-end, sequence-to-sequence probabilistic visual odometry through deep neural networks. *The International Journal of Robotics Research* **14**, 027836491773429. [Crossref]

156. Victor Greiff, Cédric R. Weber, Johannes Palme, Ulrich Bodenhofer, Enkelejda Miho, Ulrike Menzel, Sai T. Reddy. 2017. Learning the High-Dimensional Immunogenomic Features That Predict Public and Private Antibody Repertoires. *The Journal of Immunology* **199**:8, 2985-2997. [Crossref]

157. Mohamed Abdel-Nasser, Karar Mahmoud. 2017. Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Computing and Applications* **136**. . [Crossref]

158. Fabio Carrara, Andrea Esuli, Tiziano Fagni, Fabrizio Falchi, Alejandro Moreo Fernández. 2017. Picture it in your mind: generating high level visual representations from textual descriptions. *Information Retrieval Journal* **36**. . [Crossref]

159. Jin Liu, Haoliang Ren, Menglong Wu, Jin Wang, Hye-jin Kim. 2017. Multiple relations extraction among multiple entities in unstructured text. *Soft Computing* **19**. . [Crossref]

160. Xiusheng Lu, Hongxun Yao, Sicheng Zhao, Xiaoshuai Sun, Shengping Zhang. 2017. Action recognition with multi-scale trajectory-pooled 3D convolutional descriptors. *Multimedia Tools and Applications* **43**. . [Crossref]

161. Zhengqi Wen, Kehuang Li, Zhen Huang, Chin-Hui Lee, Jianhua Tao. 2017. Improving Deep Neural Network Based Speech Synthesis through Contextual Feature Parametrization and Multi-Task Learning. *Journal of Signal Processing Systems* **18**. . [Crossref]

162. Cheng-Tao Chung, Cheng-Yu Tsai, Chia-Hsiang Liu, Lin-Shan Lee. 2017. Unsupervised Iterative Deep Learning of Speech Features and Acoustic Tokens with Applications to Spoken Term Detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:10, 1914-1928. [Crossref]

163. Shuohao Li, Min Tang, Qiang Guo, Jun Lei, Jun Zhang. 2017. Deep neural network with attention model for scene text recognition. *IET Computer Vision* **11**:7, 605-612. [Crossref]

164. Yuqing Wan, Yain-Whar Si. 2017. A formal approach to chart patterns classification in financial time series. *Information Sciences* **411**, 151-175. [Crossref]

165. Zhisheng Wang, Jun Lin, Zhongfeng Wang. 2017. Accelerating Recurrent Neural Networks: A Memory-Efficient Approach. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* **25**:10, 2763-2775. [Crossref]

166. Klaus Greff, Rupesh K. Srivastava, Jan Koutnik, Bas R. Steunebrink, Jurgen Schmidhuber. 2017. LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems* **28**:10, 2222-2232. [Crossref]

167. Morten Kolbaek, Dong Yu, Zheng-Hua Tan, Jesper Jensen. 2017. Multitalker Speech Separation With Utterance-Level Permutation Invariant Training of Deep Recurrent Neural Networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:10, 1901-1913. [Crossref]

168. Lin Wu, Yang Wang, Xue Li, Junbin Gao. 2017. What-and-where to match: Deep spatially multiplicative integration networks for person re-identification. *Pattern Recognition* . [Crossref]

169. Burkhardt Funk, Matthias Rettenmeier, Tobias Lang. 2017. Deep Learning auf sequenziellen Daten als Grundlage unternehmerischer Entscheidungen. *Wirtschaftsinformatik & Management* **9**:5, 16-25. [Crossref]

170. Qin Zhang, Hui Wang, Junyu Dong, Guoqiang Zhong, Xin Sun. 2017. Prediction of Sea Surface Temperature Using Long Short-Term Memory. *IEEE Geoscience and Remote Sensing Letters* **14**:10, 1745-1749. [Crossref]

171. Long Chen, Yuhang He, Lei Fan. 2017. Let the robot tell: Describe car image with natural language via LSTM. *Pattern Recognition Letters* **98**, 75-82. [Crossref]

172. Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, Tsuhan Chen. 2017. Recent advances in convolutional neural networks. *Pattern Recognition* . [Crossref]

173. Adam Mikus, Mark Hoogendoorn, Artur Rocha, Joao Gama, Jeroen Ruwaard, Heleen Riper. 2017. Predicting short term mood developments among depressed patients using adherence and ecological momentary assessment data. *Internet Interventions* . [Crossref]

174. Lituan Wang, Lei Zhang, Zhang Yi. 2017. Trajectory Predictor by Using Recurrent Neural Networks in Visual Tracking. *IEEE Transactions on Cybernetics* **47**:10, 3172-3183. [Crossref]

175. Zufan Zhang, Yang Zou, Chenquan Gan. 2017. Textual sentiment analysis via three different attention convolutional neural networks and cross-modality consistent regression. *Neurocomputing* . [Crossref]

176. Sebastián Basterrech, Gerardo Rubino. 2017. ECHO STATE QUEUEING NETWORKS: A COMBINATION OF RESERVOIR COMPUTING AND RANDOM NEURAL NETWORKS. *Probability in the Engineering and Informational Sciences* **31**:04, 457-476. [Crossref]

177. Fei Wu, Xinyu Duan, Jun Xiao, Zhou Zhao, Siliang Tang, Yin Zhang, Yueting Zhuang. 2017. Temporal Interaction and Causal Influence in Community-Based Question Answering. *IEEE Transactions on Knowledge and Data Engineering* **29**:10, 2304-2317. [Crossref]

178. Hye Jin Kam, Ha Young Kim. 2017. Learning representations for the early detection of sepsis with deep neural networks. *Computers in Biology and Medicine* **89**, 248-255. [Crossref]

179. Yuliang Liu, Lianwen Jin, Songxuan Lai. 2017. Automatic labeling of large amounts of handwritten characters with gate-guided dynamic deep learning. *Pattern Recognition Letters* . [Crossref]

180. NhatHai Phan, Xintao Wu, Dejing Dou. 2017. Preserving differential privacy in convolutional deep belief networks. *Machine Learning* **106**:9-10, 1681-1704. [Crossref]

181. Jay H. Lee, Joohyun Shin, Matthew J. Realff. 2017. Machine learning: Overview of the recent progresses and implications for the process systems engineering field. *Computers & Chemical Engineering* . [Crossref]

182. Kangil Kim, Yun Jin, Seung-Hoon Na, Young-Kil Kim. 2017. Center-shared sliding ensemble of neural networks for syntax analysis of natural language. *Expert Systems with Applications* **83**, 215-225. [Crossref]

183. MohammadMehdi Kafashan, ShiNung Ching. 2017. Recurrent networks with soft-thresholding nonlinearities for lightweight coding. *Neural Networks* **94**, 212-219. [Crossref]

184. Xinwei He, Baoguang Shi, Xiang Bai, Gui-Song Xia, Zhaoxiang Zhang, Weisheng Dong. 2017. Image Caption Generation with Part of Speech Guidance. *Pattern Recognition Letters* . [Crossref]

185. A Rossi, F Montefoschi, A Rizzo, M Diligenti, C Festucci. 2017. Auto-Associative Recurrent Neural Networks and Long Term Dependencies in Novelty Detection for Audio Surveillance Applications. *IOP Conference Series: Materials Science and Engineering* **261**, 012009. [Crossref]

186. Youjun Li, Jiajin Huang, Haiyan Zhou, Ning Zhong. 2017. Human Emotion Recognition with Electroencephalographic Multidimensional Features by Hybrid Deep Neural Networks. *Applied Sciences* **7**:10, 1060. [Crossref]

187. Mathias Kraus, Stefan Feuerriegel. 2017. Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems* . [Crossref]

188. Mario Lassnig, Wesley Toler, Ralf Vamosi, Joaquin Bogado. 2017. Machine learning of network metrics in ATLAS Distributed Data Management. *Journal of Physics: Conference Series* **898**, 062009. [Crossref]

189. An-An Liu, Ning Xu, Yongkang Wong, Junnan Li, Yu-Ting Su, Mohan Kankanhalli. 2017. Hierarchical & multimodal video captioning: Discovering and transferring multimodal knowledge for vision to language. *Computer Vision and Image Understanding* **163**, 113-125. [Crossref]

190. Kushal Kafle, Christopher Kanan. 2017. Visual question answering: Datasets, algorithms, and future challenges. *Computer Vision and Image Understanding* **163**, 3-20. [Crossref]

191. Zhenyang Li, Kirill Gavrilyuk, Efstratios Gavves, Mihir Jain, Cees G.M. Snoek. 2017. VideoLSTM convolves, attends and flows for action recognition. *Computer Vision and Image Understanding* . [Crossref]

192. Nikodem Rybak, Maureen Hassall, Kourosh Parsa, Daniel J. Angus. New real-time methods for operator situational awareness retrieval and higher process safety in the control room 1-7. [Crossref]

193. Aditya Tewari, Bertram Taetz, Frederic Grandidier, Didier Stricker. [POSTER] A Probabilistic Combination of CNN and RNN Estimates for Hand Gesture Based Interaction in Car 1-6. [Crossref]

194. Mauro Dragoni, Giulio Petrucci. 2017. A Neural Word Embeddings Approach for Multi-Domain Sentiment Analysis. *IEEE Transactions on Affective Computing* **8**:4, 457-470. [Crossref]

195. Chenglin Xu, Lei Xie, Xiong Xiao. 2017. A Bidirectional LSTM Approach with Word Embeddings for Sentence Boundary Detection. *Journal of Signal Processing Systems* **32**. . [Crossref]

196. Jian Kang, Wei-Qiang Zhang, Wei-Wei Liu, Jia Liu, Michael T. Johnson. 2017. Lattice Based Transcription Loss for End-to-End Speech Recognition. *Journal of Signal Processing Systems* **20**. . [Crossref]

197. Yibin Zheng, Ya Li, Zhengqi Wen, Bin Liu, Jianhua Tao. 2017. Investigating Deep Neural Network Adaptation for Generating Exclamatory and Interrogative Speech in Mandarin. *Journal of Signal Processing Systems* **13**. . [Crossref]

198. Dorien Herremans, Ching-Hua Chuan, Elaine Chew. 2017. A Functional Taxonomy of Music Generation Systems. *ACM Computing Surveys* **50**:5, 1-30. [Crossref]

199. Rhys Heffernan, Yuedong Yang, Kuldip Paliwal, Yaoqi Zhou. 2017. Capturing non-local interactions by long short-term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers and solvent accessibility. *Bioinformatics* **33**:18, 2842-2849. [Crossref]

200. Umut Güçlü, Marcel van Gerven. Probing Human Brain Function with Artificial Neural Networks 413-423. [Crossref]

201. Asif Salekin, Zeya Chen, Mohsin Y. Ahmed, John Lach, Donna Metz, Kayla De La Haye, Brooke Bell, John A. Stankovic. 2017. Distant Emotion Recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **1**:3, 1-25. [Crossref]

202. Ákos Kádár, Grzegorz Chrupała, Afra Alishahi. Representation of Linguistic Form and Function in Recurrent Neural Networks. *Computational Linguistics*, ahead of print761-780. [Abstract] [PDF] [PDF Plus]

203. Peyman Passban, Qun Liu, Andy Way. 2017. Translating Low-Resource Languages by Vocabulary Adaptation from Close Counterparts. *ACM Transactions on Asian and Low-Resource Language Information Processing* **16**:4, 1-14. [Crossref]

204. REKIA KADARI, YU ZHANG, WEINAN ZHANG, TING LIU. 2017. CCG supertagging with bidirectional long short-term memory networks. *Natural Language Engineering* **8**, 1-14. [Crossref]

205. Xi Xiao, Shaofeng Zhang, Francesco Mercaldo, Guangwu Hu, Arun Kumar Sangaiah. 2017. Android malware detection based on system call sequences and LSTM. *Multimedia Tools and Applications* **2**. . [Crossref]

206. Waseem Rawat, Zenghui Wang. 2017. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation* **29**:9, 2352-2449. [Abstract] [Full Text] [PDF] [PDF Plus]

207. Francisco Guzmán, Shafiq Joty, Lluís Màrquez, Preslav Nakov. 2017. Machine translation evaluation with neural networks. *Computer Speech & Language* **45**, 180-200. [Crossref]

208. Maciej Wielgosz, Andrzej Skocze?, Matej Mertik. 2017. Using LSTM recurrent neural networks for monitoring the LHC superconducting magnets. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **867**, 40-50. [Crossref]

209. Luciano Sánchez, Inés Couso, Cecilio Blanco. 2017. A class of Monotone Fuzzy rule-based Wiener systems with an application to Li-ion battery modelling. *Engineering Applications of Artificial Intelligence* **64**, 367-377. [Crossref]

210. Basura Fernando, Stephen Gould. 2017. Discriminatively Learned Hierarchical Rank Pooling Networks. *International Journal of Computer Vision* **124**:3, 335-355. [Crossref]

211. Álvaro Peris, Miguel Domingo, Francisco Casacuberta. 2017. Interactive neural machine translation. *Computer Speech & Language* **45**, 201-220. [Crossref]

212. Heeyoul Choi, Kyunghyun Cho, Yoshua Bengio. 2017. Context-dependent word representation for neural machine translation. *Computer Speech & Language* **45**, 149-160. [Crossref]

213. Antonio Jimeno Yepes. 2017. Word embeddings and recurrent neural networks based on Long-Short Term Memory nodes in supervised biomedical word sense disambiguation. *Journal of Biomedical Informatics* **73**, 137-147. [Crossref]

214. Linchao Zhu, Zhongwen Xu, Yi Yang, Alexander G. Hauptmann. 2017. Uncovering the Temporal Context for Video Question Answering. *International Journal of Computer Vision* **124**:3, 409-421. [Crossref]

215. Caglar Gulcehre, Orhan Firat, Kelvin Xu, Kyunghyun Cho, Yoshua Bengio. 2017. On integrating a language model into neural machine translation. *Computer Speech & Language* **45**, 137-148. [Crossref]

216. Lianli Gao, Zhao Guo, Hanwang Zhang, Xing Xu, Heng Tao Shen. 2017. Video Captioning With Attention-Based LSTM and Semantic Consistency. *IEEE Transactions on Multimedia* **19**:9, 2045-2055. [Crossref]

217. , , . 2017. A Long-Short Term Memory Recurrent Neural Network Based Reinforcement Learning Controller for Office Heating Ventilation and Air Conditioning Systems. *Processes* **5**:3, 46. [Crossref]

218. Weining Lu, Yu Cheng, Cao Xiao, Shiyu Chang, Shuai Huang, Bin Liang, Thomas Huang. 2017. Unsupervised Sequential Outlier Detection With Deep Architectures. *IEEE Transactions on Image Processing* **26**:9, 4321-4330. [Crossref]

219. Jing Han, Zixing Zhang, Nicholas Cummins, Fabien Ringeval, Björn Schuller. 2017. Strength modelling for real-worldautomatic continuous affect recognition from audiovisual signals. *Image and Vision Computing* **65**, 76-86. [Crossref]

220. Xiaocong Wei, Hongfei Lin, Liang Yang, Yuhai Yu. 2017. A Convolution-LSTM-Based Deep Neural Network for Cross-Domain MOOC Forum Post Classification. *Information* **8**:3, 92. [Crossref]

221. Sundaram Laksshman, Rajendra Rana Bhat, Vivek Viswanath, Xiaolin Li. 2017. DeepBipolar: Identifying genomic mutations for bipolar disorder via deep learning. *Human Mutation* **38**:9, 1217-1224. [Crossref]

222. Duncan Taylor, Ash Harrison, David Powers. 2017. An artificial neural network system to identify alleles in reference electropherograms. *Forensic Science International: Genetics* **30**, 114-126. [Crossref]

223. Orhan Firat, Kyunghyun Cho, Baskaran Sankaran, Fatos T. Yarman Vural, Yoshua Bengio. 2017. Multi-way, multilingual neural machine translation. *Computer Speech & Language* **45**, 236-252. [Crossref]

224. Yaming Sun, Zhenzhou Ji, Lei Lin, Duyu Tang, Xiaolong Wang. 2017. Entity disambiguation with decomposable neural networks. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **7**:5, e1215. [Crossref]

225. Chunyi Yue, Hanqiang Cao, Kun Xiong, Anqi Cui, Haocheng Qin, Ming Li. 2017. Enhanced question understanding with dynamic memory networks for textual question answering. *Expert Systems with Applications* **80**, 39-45. [Crossref]

226. William Gu, Gerald Seet, Nadia Magnenat-Thalmanna. 2017. Perception-Link Behavior Model: Supporting a Novel Operator Interface for a Customizable Anthropomorphic Telepresence Robot. *Robotics* **6**:3, 16. [Crossref]

227. Eric Aislan Antonelo, Carlos Alberto Flesch, Filipe Schmitz. 2017. Reservoir computing for detection of steady state in performance tests of compressors. *Neurocomputing* . [Crossref]

228. Frank E. Curtis, Katya Scheinberg. Optimization Methods for Supervised Machine Learning: From Linear Models to Deep Learning 89-113. [Crossref]

229. Lemao Liu, Atsushi Fujita, Masao Utiyama, Andrew Finch, Eiichiro Sumita. 2017. Translation Quality Estimation Using Only Bilingual Corpora. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:9, 1458-1468. [Crossref]

230. Xiaochuan Sun, Tao Li, Yingqi Li, Qun Li, Yue Huang, Jiayu Liu. 2017. Recurrent neural system with minimum complexity: A deep learning perspective. *Neurocomputing* . [Crossref]

231. Haonan Yu, Jeffrey Mark Siskind. 2017. Sentence Directed Video Object Codiscovery. *International Journal of Computer Vision* **124**:3, 312-334. [Crossref]

232. Na Zhao, Hanwang Zhang, Richang Hong, Meng Wang, Tat-Seng Chua. 2017. VideoWhisper: Toward Discriminative Unsupervised Video Feature Learning With Attention-Based Recurrent Neural Networks. *IEEE Transactions on Multimedia* **19**:9, 2080-2092. [Crossref]

233. Yiwei Zhang, Chao Wang, Lei Gong, Yuntao Lu, Fan Sun, Chongchong Xu, Xi Li, Xuehai Zhou. A Power-Efficient Accelerator Based on FPGAs for LSTM Network 629-630. [Crossref]

234. Seong Tae Kim, Yeoreum Choi, Yong Man Ro. Multi-Scale Facial Scanning via Spatial Lstm for Latent Facial Feature Representation 1-5. [Crossref]

235. Kristijan Armeni, Roel M. Willems, Stefan L. Frank. 2017. Probabilistic language models in cognitive neuroscience: Promises and pitfalls. *Neuroscience & Biobehavioral Reviews* . [Crossref]

236. Chris Cummins, Pavlos Petoumenos, Zheng Wang, Hugh Leather. End-to-End Deep Learning of Optimization Heuristics 219-232. [Crossref]

237. Chixiao Chen, Hongwei Ding, Huwan Peng, Haozhe Zhu, Rui Ma, Peiyong Zhang, Xiaolang Yan, Yu Wang, Mingyu Wang, Hao Min, Richard C.-J Shi. OCEAN: An on-chip incremental-learning enhanced processor with gated recurrent neural network accelerators 259-262. [Crossref]

238. Abhinav Valada, Wolfram Burgard. 2017. Deep spatiotemporal models for robust proprioceptive terrain classification. *The International Journal of Robotics Research* **32**, 027836491772706. [Crossref]

239. Zhibin Yu, Dennis S. Moirangthem, Minho Lee. 2017. Continuous Timescale Long-Short Term Memory Neural Network for Human Intent Understanding. *Frontiers in Neurorobotics* **11**. . [Crossref]

240. Brandon Reagen, Robert Adolf, Paul Whatmough, Gu-Yeon Wei, David Brooks. 2017. Deep Learning for Computer Architects. *Synthesis Lectures on Computer Architecture* **12**:4, 1-123. [Crossref]

241. Saad Bin Ahmed, Saeeda Naz, Salahuddin Swati, Muhammad Imran Razzak. 2017. Handwritten Urdu character recognition using one-dimensional BLSTM classifier. *Neural Computing and Applications* **29**. . [Crossref]

242. Harry A. Pierson, Michael S. Gashler. 2017. Deep learning in robotics: a review of recent research. *Advanced Robotics* **31**:16, 821-835. [Crossref]

243. Jure Sokolic, Raja Giryes, Guillermo Sapiro, Miguel R. D. Rodrigues. 2017. Robust Large Margin Deep Neural Networks. *IEEE Transactions on Signal Processing* **65**:16, 4265-4280. [Crossref]

244. Z. Wang, D. Xiao, F. Fang, R. Govindan, C. C. Pain, Y. Guo. 2017. Model identification of reduced order fluid dynamics systems using deep learning. *International Journal for Numerical Methods in Fluids* **54**. . [Crossref]

245. Du-Xin Liu, Xinyu Wu, Wenbin Du, Can Wang, Chunjie Chen, Tiantian Xu. 2017. Deep Spatial-Temporal Model for rehabilitation gait: optimal trajectory generation for knee joint of lower-limb exoskeleton. *Assembly Automation* **37**:3, 369-378. [Crossref]

246. Ilya S. Ardakani, Koichi Hashimoto. Encoding bird's trajectory using Recurrent Neural Networks 1644-1649. [Crossref]

247. Euisok Chung, Jeon Gue Park. 2017. Sentence-Chain Based Seq2seq Model for Corpus Expansion. *ETRI Journal* **39**:4, 455-466. [Crossref]

248. Enrico Prati. 2017. Quantum neuromorphic hardware for quantum artificial intelligence. *Journal of Physics: Conference Series* **880**, 012018. [Crossref]

249. Nina Dethlefs. 2017. Domain Transfer for Deep Natural Language Generation from Abstract Meaning Representations. *IEEE Computational Intelligence Magazine* **12**:3, 18-28. [Crossref]

250. Yue Gu, Xinyu Li, Shuhong Chen, Hunagcan Li, Richard A. Farneth, Ivan Marsic, Randall S. Burd. Language-Based Process Phase Detection in the Trauma Resuscitation 239-247. [Crossref]

251. Yong Zhuang, Matthew Almeida, Melissa Morabito, Wei Ding. Crime Hot Spot Forecasting: A Recurrent Model with Spatial and Temporal Information 143-150. [Crossref]

252. Truc Le, Giang Bui, Ye Duan. 2017. A multi-view recurrent neural network for 3D mesh segmentation. *Computers & Graphics* **66**, 103-112. [Crossref]

253. Dashan Guo, Wei Li, Xiangzhong Fang. 2017. Capturing Temporal Structures for Video Captioning by Spatio-temporal Contexts and Channel Attention Mechanism. *Neural Processing Letters* **46**:1, 313-328. [Crossref]

254. Mingai Li, Wei Zhu, Meng Zhang, Yanjun Sun, Zhe Wang. The novel recognition method with Optimal Wavelet Packet and LSTM based Recurrent Neural Network 584-589. [Crossref]

255. Ngoc Hieu Tran, Xianglilan Zhang, Lei Xin, Baozhen Shan, Ming Li. 2017. De novo peptide sequencing by deep learning. *Proceedings of the National Academy of Sciences* **114**:31, 8247-8252. [Crossref]

256. Bo-Sheng Lin, Chuin-Mu Wang, Cheng-Ning Yu. The establishment of human-computer interaction based on Word2Vec 1698-1703. [Crossref]

257. Joerg Evermann, Jana-Rebecca Rehse, Peter Fettke. 2017. Predicting process behaviour using deep learning. *Decision Support Systems* **100**, 129-140. [Crossref]

258. Jon Ander Gómez, Juan Arévalo, Roberto Paredes, Jordi Nin. 2017. End-to-end neural network architecture for fraud scoring in card payments. *Pattern Recognition Letters* . [Crossref]

259. Keisuke Fujii, Kohei Nakajima. 2017. Harnessing Disordered-Ensemble Quantum Dynamics for Machine Learning. *Physical Review Applied* **8**:2. . [Crossref]

260. Huiting Zheng, Jiabin Yuan, Long Chen. 2017. Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. *Energies* **10**:8, 1168. [Crossref]

261. Haytham M. Fayek, Margaret Lech, Lawrence Cavedon. 2017. Evaluating deep learning architectures for Speech Emotion Recognition. *Neural Networks* **92**, 60-68. [Crossref]

262. Salvatore Romeo, Giovanni Da San Martino, Yonatan Belinkov, Alberto Barrón-Cedeño, Mohamed Eldesouki, Kareem Darwish, Hamdy Mubarak, James Glass, Alessandro Moschitti. 2017. Language processing and learning models for community question answering in Arabic. *Information Processing & Management* . [Crossref]

263. Juan Sebastian Olier, Emilia Barakova, Carlo Regazzoni, Matthias Rauterberg. 2017. Re-framing the characteristics of concepts and their relation to learning and cognition in artificial agents. *Cognitive Systems Research* **44**, 50-68. [Crossref]

264. Guangwu Qian, Lei Zhang. 2017. A simple feedforward convolutional conceptor neural network for classification. *Applied Soft Computing* . [Crossref]

265. Sabato Marco Siniscalchi, Valerio Mario Salerno. 2017. Adaptation to New Microphones Using Artificial Neural Networks With Trainable Activation Functions. *IEEE Transactions on Neural Networks and Learning Systems* **28**:8, 1959-1965. [Crossref]

266. Ramakanth Kavuluru, Anthony Rios, Tung Tran. Extracting Drug-Drug Interactions with Word and Character-Level Recurrent Neural Networks 5-12. [Crossref]

267. Yuzhen Lu, Fathi M. Salem. Simplified gating in long short-term memory (LSTM) recurrent neural networks 1601-1604. [Crossref]

268. Rahul Dey, Fathi M. Salemt. Gate-variants of Gated Recurrent Unit (GRU) neural networks 1597-1600. [Crossref]

269. Federico Colangelo, Federica Battisti, Marco Carli, Alessandro Neri, Francesco Calabro. Enhancing audio surveillance with hierarchical recurrent neural networks 1-6. [Crossref]

270. F. Scheidegger, L. Cavigelli, M. Schaffner, A. C. I. Malossi, C. Bekas, L. Benini. Impact of temporal subsampling on accuracy and performance in practical video classification 996-1000. [Crossref]

271. Wenliang Shao, Chunhong Zhang, Tingting Sun, Hang Li, Yang Ji, Xiaofeng Qiu. Hierarchical Bidirectional Long Short-Term Memory Networks for Chinese Messaging Spam Filtering 158-164. [Crossref]

272. Bartosz Kostka, Jaroslaw Kwiecieli, Jakub Kowalski, Pawel Rychlikowski. Text-based adventures of the golovin AI agent 181-188. [Crossref]

273. Avraam Tsantekidis, Nikolaos Passalis, Anastasios Tefas, Juho Kanniainen, Moncef Gabbouj, Alexandros Iosifidis. Using deep learning to detect price change indications in financial markets 2511-2515. [Crossref]

274. Takayuki Kawashima, Yasutomo Kawanishi, Ichiro Ide, Hiroshi Murase, Daisuke Deguchi, Tomoyoshi Aizawa, Masato Kawade. Action recognition from extremely low-resolution thermal image sequence 1-6. [Crossref]

275. Swathikiran Sudhakaran, Oswald Lanz. Learning to detect violent videos using convolutional long short-term memory 1-6. [Crossref]

276. Jian Zheng, Bin Tong, Yoshiyasu Takahashi, Yuichi Kobayashi, Tatsuhiro Sato, Iwao Tanuma, Anshuman Sahu, Ravigopal Vennelakanti. Long short-term memory model for predicting productivity of drilling space units 83-88. [Crossref]

277. Anh Minh Truong, Atsuo Yoshitaka. Structured LSTM for human-object interaction detection and anticipation 1-6. [Crossref]

278. Naomichi Nakamura, Hitoshi Isahara. Effect of linguistic information in neural machine translation 1-6. [Crossref]

279. Tomoki Nishimura, Tomoyosi Akiba. Addressing unknown word problem for neural machine translation using distributee representations of words as input features 1-6. [Crossref]

280. Stefan Braun, Daniel Neil, Shih-Chii Liu. A curriculum learning method for improved noise robustness in automatic speech recognition 548-552. [Crossref]

281. Tong Sun, Jia Wang, Pengfei Zhang, Yu Cao, Benyuan Liu, Degang Wang. Predicting Stock Price Returns Using Microblog Sentiment for Chinese Stock Market 87-96. [Crossref]

282. Sarah Taylor, Taehwan Kim, Yisong Yue, Moshe Mahler, James Krahe, Anastasio Garcia Rodriguez, Jessica Hodgins, Iain Matthews. 2017. A deep learning approach for generalized speech animation. *ACM Transactions on Graphics* **36**:4, 1-11. [Crossref]

283. Supasorn Suwajanakorn, Steven M. Seitz, Ira Kemelmacher-Shlizerman. 2017. Synthesizing Obama. *ACM Transactions on Graphics* **36**:4, 1-13. [Crossref]

284. Maryam Habibi, Leon Weber, Mariana Neves, David Luis Wiegandt, Ulf Leser. 2017. Deep learning with word embeddings improves biomedical named entity recognition. *Bioinformatics* **33**:14, i37-i48. [Crossref]

285. Xu Min, Wanwen Zeng, Ning Chen, Ting Chen, Rui Jiang. 2017. Chromatin accessibility prediction via convolutional long short-term memory networks with k-mer embedding. *Bioinformatics* **33**:14, i92-i101. [Crossref]

286. Wei Bao, Jun Yue, Yulei Rao. 2017. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLOS ONE* **12**:7, e0180944. [Crossref]

287. Jordan R. Ubbens, Ian Stavness. 2017. Deep Plant Phenomics: A Deep Learning Platform for Complex Plant Phenotyping Tasks. *Frontiers in Plant Science* **8**. . [Crossref]

288. Seungwook Kim, Daeyoung Choi, Eunjung Lee, Wonjong Rhee. 2017. Churn prediction of mobile and online casual games using play log data. *PLOS ONE* **12**:7, e0180735. [Crossref]

289. Haiqing Ren, Weiqiang Wang, Ke Lu, Jianshe Zhou, Qiuchen Yuan. An end-to-end recognizer for in-air handwritten Chinese characters based on a new recurrent neural networks 841-846. [Crossref]

290. Zhang Miao, Lin Zheqi, Jia Peiyuan, Shen Yi. Change detection method for hyperspectral image with sequential time series inputs 10819-10822. [Crossref]

291. Yuan Luo. 2017. Recurrent Neural Networks for Classifying Relations in Clinical Notes. *Journal of Biomedical Informatics* . [Crossref]

292. Feng Gag, Yihang Lou, Yan Bai, Shiqi Wang, Tiejun Huang, Ling-Yu Duan. Improving object detection with region similarity learning 1488-1493. [Crossref]

293. Alican Bozkurt, Trevor Gale, Kivanc Kose, Christi Alessi-Fox, Dana H. Brooks, Milind Rajadhyaksha, Jennifer Dy. Delineation of Skin Strata in Reflectance Confocal Microscopy Images with Recurrent Convolutional Networks 777-785. [Crossref]

294. Takamasa Tsunoda, Yasuhiro Komori, Masakazu Matsugu, Tatsuya Harada. Football Action Recognition Using Hierarchical LSTM 155-163. [Crossref]

295. Senlin Zhang, Siyang Liu, Meiqin Liu. Natural language inference using LSTM model with sentence fusion 11081-11085. [Crossref]

296. Zaifa Chen, Yancheng Liu, Siyuan Liu. Mechanical state prediction based on LSTM neural netwok 3876-3881. [Crossref]

297. Pavol Harar, Jesus B. Alonso-Hernandezy, Jiri Mekyska, Zoltan Galaz, Radim Burget, Zdenek Smekal. Voice Pathology Detection Using Deep Learning: a Preliminary Study 1-4. [Crossref]

298. Zhuo-Ran Liu, Yang Liu. 2017. Exploiting Unlabeled Data for Neural Grammatical Error Detection. *Journal of Computer Science and Technology* **32**:4, 758-767. [Crossref]

299. Zixing Zhang, Nicholas Cummins, Bjoern Schuller. 2017. Advanced Data Exploitation in Speech Analysis: An overview. *IEEE Signal Processing Magazine* **34**:4, 107-129. [Crossref]

300. Lichao Mou, Pedram Ghamisi, Xiao Xiang Zhu. 2017. Deep Recurrent Neural Networks for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing* **55**:7, 3639-3655. [Crossref]

301. Shu-Chang Zhou, Yu-Zhi Wang, He Wen, Qin-Yao He, Yu-Heng Zou. 2017. Balanced Quantization: An Effective and Efficient Approach to Quantized Neural Networks. *Journal of Computer Science and Technology* **32**:4, 667-682. [Crossref]

302. Svati Dhamija, Terrance E. Boult. Exploring Contextual Engagement for Trauma Recovery 2267-2277. [Crossref]

303. Qiurui Wang, Chun Yuan, Zhihui Lin. Learning attentional recurrent neural network for visual tracking 1237-1242. [Crossref]

304. Henry Lin, Max Tegmark. 2017. Critical Behavior in Physics and Probabilistic Formal Languages. *Entropy* **19**:7, 299. [Crossref]

305. Haiyang Yu, Zhihai Wu, Shuqin Wang, Yunpeng Wang, Xiaolei Ma. 2017. Spatiotemporal Recurrent Convolutional Networks for Traffic Prediction in Transportation Networks. *Sensors* **17**:7, 1501. [Crossref]

306. Kaustuv Mishra, Rachana Sathish, Debdoot Sheet. Learning Latent Temporal Connectionism of Deep Residual Visual Abstractions for Identifying Surgical Tools in Laparoscopy Procedures 2233-2240. [Crossref]

307. K. Seeliger, M. Fritsche, U. Güçlü, S. Schoenmakers, J.-M. Schoffelen, S.E. Bosch, M.A.J. van Gerven. 2017. Convolutional neural network-based encoding and decoding of visual object recognition in space and time. *NeuroImage* . [Crossref]

308. Xudong Li, Mao Ye, Yiguang Liu, Feng Zhang, Dan Liu, Song Tang. 2017. Accurate object detection using memory-based models in surveillance scenes. *Pattern Recognition* **67**, 73-84. [Crossref]

309. Behzad Hasani, Mohammad H. Mahoor. Facial Expression Recognition Using Enhanced Deep 3D Convolutional Neural Networks 2278-2288. [Crossref]

310. G. Saon, M. Picheny. 2017. Recent advances in conversational speech recognition using convolutional and recurrent neural networks. *IBM Journal of Research and Development* **61**:4, 1:1-1:10. [Crossref]

311. Peiye Liu, Wu Liu, Huadong Ma. Weighted sequence loss based spatial-temporal deep learning framework for human body orientation estimation 97-102. [Crossref]

312. Shimon Edelman. 2017. Language and other complex behaviors: Unifying characteristics, computational models, neural mechanisms. *Language Sciences* **62**, 91-123. [Crossref]

313. Na Zhao, Hanwang Zhang, Mingxing Zhang, Richang Hong, Meng Wang, Tat-seng Chua. VIDEOWHISPER: Towards unsupervised learning of discriminative features of videos with RNN 277-282. [Crossref]

314. Xudong Li, Mao Ye, Yiguang Liu, Ce Zhu. Memory-based pedestrian detection through sequence learning 1129-1134. [Crossref]

315. Gil Keren, Tobias Kirschstein, Erik Marchi, Fabien Ringeval, Bjorn Schuller. End-to-end learning for dimensional emotion recognition from physiological signals 985-990. [Crossref]

316. Shi-Qi Shen, Yang Liu, Mao-Song Sun. 2017. Optimizing Non-Decomposable Evaluation Metrics for Neural Machine Translation. *Journal of Computer Science and Technology* **32**:4, 796-804. [Crossref]

317. Tae Soo Kim, Austin Reiter. Interpretable 3D Human Action Analysis with Temporal Convolutional Networks 1623-1631. [Crossref]

318. Kai-Yuan Cui, Peng-Jie Ren, Zhu-Min Chen, Tao Lian, Jun Ma. 2017. Relation Enhanced Neural Model for Type Classification of Entity Mentions with a Fine-Grained Taxonomy. *Journal of Computer Science and Technology* **32**:4, 814-827. [Crossref]

319. Zhao Lin, Kefeng Ji, Miao Kang, Xiangguang Leng, Huanxin Zou. 2017. Deep Convolutional Highway Unit Network for SAR Target Classification With Limited Labeled Training Data. *IEEE Geoscience and Remote Sensing Letters* **14**:7, 1091-1095. [Crossref]

320. Ryan G. Hefron, Brett J. Borghetti, James C. Christensen, Christine M. Schubert Kabban. 2017. Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation. *Pattern Recognition Letters* **94**, 96-104. [Crossref]

321. Yemin Shi, Yonghong Tian, Yaowei Wang, Tiejun Huang. 2017. Sequential Deep Trajectory Descriptor for Action Recognition With Three-Stream CNN. *IEEE Transactions on Multimedia* **19**:7, 1510-1520. [Crossref]

322. Moumita Roy Tora, Jianhui Chen, James J. Little. Classification of Puck Possession Events in Ice Hockey 147-154. [Crossref]

323. Yuesong Shen, Claire-Heiene Demarty, Ngoc Q. K. Duong. Deep learning for multimodal-based video interestingness prediction 1003-1008. [Crossref]

324. Supriyo Chakraborty, Alun Preece, Moustafa Alzantot, Tianwei Xing, Dave Braines, Mani Srivastava. Deep learning for situational understanding 1-8. [Crossref]

325. Chunchun Yang, Zhiheng Li, Bin Liang, Qiangqiang Cui, Weining Lu, Jun Yang. A novel fusion strategy for failure prognostic of proton exchange membrane fuel cell stack 7070-7076. [Crossref]

326. Hamid Reza Hassanzadeh, Ying Sha, May D. Wang. DeepDeath: Learning to predict the underlying cause of death with Big Data 3373-3376. [Crossref]

327. Liang Lin, Lili Huang, Tianshui Chen, Yukang Gan, Hui Cheng. Knowledge-guided recurrent neural network learning for task-oriented action prediction 625-630. [Crossref]

328. Mandy Korpusik, James Glass. 2017. Spoken Language Understanding for a Nutrition Dialogue System. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:7, 1450-1461. [Crossref]

329. Bob D. de Vos, Jelmer M. Wolterink, Pim A. de Jong, Tim Leiner, Max A. Viergever, Ivana Isgum. 2017. ConvNet-Based Localization of Anatomical Structures in 3-D Medical Images. *IEEE Transactions on Medical Imaging* **36**:7, 1470-1481. [Crossref]

330. Shengtao Xiao, Jianshu Li, Yunpeng Chen, Zhecan Wang, Jiashi Feng, Shuicheng Yan, Ashraf Kassim. 3D-Assisted Coarse-to-Fine Extreme-Pose Facial Landmark Detection 2060-2068. [Crossref]

331. Reid Pryzant, Stefano Ermon, David Lobell. Monitoring Ethiopian Wheat Fungus with Satellite Imagery and Deep Feature Learning 1524-1532. [Crossref]

332. Qiaoyun Wang, He Huang. Learning of recurrent convolutional neural networks with applications in pattern recognition 4135-4139. [Crossref]

333. Hao Hu, Zhaowen Wang, Joon-Young Lee, Zhe Lin, Guo-Jun Qi. Temporal Domain Neural Encoder for Video Representation Learning 2192-2199. [Crossref]

334. Zengjian Liu, Ming Yang, Xiaolong Wang, Qingcai Chen, Buzhou Tang, Zhe Wang, Hua Xu. 2017. Entity recognition from clinical texts via recurrent neural network. *BMC Medical Informatics and Decision Making* **17**:S2. . [Crossref]

335. Michele Mancini, Gabriele Costante, Paolo Valigi, Thomas A. Ciarfuglia, Jeffrey Delmerico, Davide Scaramuzza. 2017. Toward Domain Independence for Learning-Based Monocular Depth Estimation. *IEEE Robotics and Automation Letters* **2**:3, 1778-1785. [Crossref]

336. Fei Hu, Li Li, Zi-Li Zhang, Jing-Yuan Wang, Xiao-Fei Xu. 2017. Emphasizing Essential Words for Sentiment Classification Based on Recurrent Neural Networks. *Journal of Computer Science and Technology* **32**:4, 785-795. [Crossref]

337. Jan Jakubik, Halina Kwasnicka. Music emotion analysis using semantic embedding recurrent neural networks 271-276. [Crossref]

338. Frank P.-W. Lo, Charles X.-T. Li, Jiankun Wang, Jiyu Cheng, Max Q.-H. Meng. Continuous systolic and diastolic blood pressure estimation utilizing long short-term memory network 1853-1856. [Crossref]

339. Liang Lu, Steve Renals. 2017. Small-Footprint Highway Deep Neural Networks for Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:7, 1502-1511. [Crossref]

340. Daniel Lopez Martinez, Ognjen Rudovic, Rosalind Picard. Personalized Automatic Estimation of Self-Reported Pain Intensity from Facial Expressions 2318-2327. [Crossref]

341. Samira Ebrahimi Kahou, Vincent Michalski, Roland Memisevic, Christopher Pal, Pascal Vincent. RATM: Recurrent Attentive Tracking Model 1613-1622. [Crossref]

342. Hao Liu, Jiashi Feng, Meibin Qi, Jianguo Jiang, Shuicheng Yan. 2017. End-to-End Comparative Attention Networks for Person Re-Identification. *IEEE Transactions on Image Processing* **26**:7, 3492-3506. [Crossref]

343. Jung-Hwan Park, Seong-Joon Yoo, Kyung-Joong Kim, Yeong-Hyeon Gu, Keon-Hoon Lee, U-Hyon Son. PM10 density forecast model using long short term memory 576-581. [Crossref]

344. Martin Stepnicka, Michal Burda. On the results and observations of the time series forecasting competition CIF 2016 1-6. [Crossref]

345. Mustafa Ridha Mezaal, Biswajeet Pradhan, Maher Ibrahim Sameen, Helmi Zulhaidi Mohd Shafri, Zainuddin Md Yusoff. 2017. Optimized Neural Architecture for Automatic Landslide Detection from High-Resolution Airborne Laser Scanning Data. *Applied Sciences* **7**:7, 730. [Crossref]

346. Yue Qiu, Feng Xiao, Haibin Shen. Elimination of power line interference from ECG signals using recurrent neural networks 2296-2299. [Crossref]

347. K. Audhkhasi, A. Rosenberg, G. Saon, A. Sethy, B. Ramabhadran, S. Chen, M. Picheny. 2017. Recent progress in deep end-to-end models for spoken language processing. *IBM Journal of Research and Development* **61**:4, 2:1-2:10. [Crossref]

348. Tao Zhou, Muhao Chen, Jie Yu, Demetri Terzopoulos. Attention-Based Natural Language Person Retrieval 27-34. [Crossref]

349. Behzad Hasani, Mohammad H. Mahoor. Facial Affect Estimation in the Wild Using Deep Residual and Convolutional Networks 1955-1962. [Crossref]

350. Ayana, Shi-Qi Shen, Yan-Kai Lin, Cun-Chao Tu, Yu Zhao, Zhi-Yuan Liu, Mao-Song Sun. 2017. Recent Advances on Neural Headline Generation. *Journal of Computer Science and Technology* **32**:4, 768-784. [Crossref]

351. John Grisham, Sagar Samtani, Mark Patton, Hsinchun Chen. Identifying mobile malware and key threat actors in online hacker forums for proactive cyber threat intelligence 13-18. [Crossref]

352. Arnav Kumar Jain, Abhinav Agarwalla, Kumar Krishna Agrawal, Pabitra Mitra. Recurrent Memory Addressing for Describing Videos 2200-2207. [Crossref]

353. Yun Ren, Zulin Wang, Mai Xu, Haoyu Dong, Shengxi Li. Learning Dynamic GMM for Attention Distribution on Single-Face Videos 1632-1641. [Crossref]

354. Marc RuBwurm, Marco Korner. Temporal Vegetation Modelling Using Long Short-Term Memory Networks for Crop Identification from Medium-Resolution Multi-spectral Satellite Images 1496-1504. [Crossref]

355. Geonmo Gu, Seong Tae Kim, Yong Man Ro. Adaptive attention fusion network for visual question answering 997-1002. [Crossref]

356. Lin-Er Yang, Mao-Song Sun, Yong Cheng, Jia-Cheng Zhang, Zheng-Hao Liu, Huan-Bo Luan, Yang Liu. 2017. Neural Parse Combination. *Journal of Computer Science and Technology* **32**:4, 749-757. [Crossref]

357. Hongsong Wang, Liang Wang. Learning robust representations using recurrent neural networks for skeleton based action classification and detection 591-596. [Crossref]

358. Anne Cocos, Alexander G Fiks, Aaron J Masino. 2017. Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. *Journal of the American Medical Informatics Association* **24**:4, 813-821. [Crossref]

359. Xin Li, Yanghui Rao, Haoran Xie, Xuebo Liu, Tak-Lam Wong, Fu Lee Wang. 2017. Social emotion classification based on noise-aware training. *Data & Knowledge Engineering* . [Crossref]

360. Shuo Hong Wang, Hai Feng Su, Xi En Cheng, Ye Liu, Aike Quo, Yan Qiu Chen. Tracking the 3D position and orientation of flying swarms with learned kinematic pattern using LSTM network 1225-1230. [Crossref]

361. Ting Bai, Hong-Jian Dou, Wayne Xin Zhao, Ding-Yi Yang, Ji-Rong Wen. 2017. An Experimental Study of Text Representation Methods for Cross-Site Purchase Preference Prediction Using the Social Text Data. *Journal of Computer Science and Technology* **32**:4, 828-842. [Crossref]

362. Po-Jen Chen, Jian-Jiun Ding, Hung-Wei Hsu, Chien-Yao Wang, Jia-Ching Wang. Improved convolutional neural network based scene classification using long short-term memory and label relations 429-434. [Crossref]

363. Demis Hassabis, Dharshan Kumaran, Christopher Summerfield, Matthew Botvinick. 2017. Neuroscience-Inspired Artificial Intelligence. *Neuron* **95**:2, 245-258. [Crossref]

364. Xiangbo Shu, Jinhui Tang, Guo-Jun Qi, Yan Song, Zechao Li, Liyan Zhang. Concurrence-Aware Long Short-Term Sub-Memories for Person-Person Action Recognition 2176-2183. [Crossref]

365. Yossi Adi, Einat Kermany, Yonatan Belinkov, Ofer Lavi, Yoav Goldberg. 2017. Analysis of sentence embedding models using prediction tasks in natural language processing. *IBM Journal of Research and Development* **61**:4, 3:1-3:9. [Crossref]

366. Silvia Vinyes Mora, William J. Knottenbelt. Deep Learning for Domain-Specific Action Recognition in Tennis 170-178. [Crossref]

367. Hai X. Pham, Samuel Cheung, Vladimir Pavlovic. Speech-Driven 3D Facial Animation with Implicit Emotional Awareness: A Deep Learning Approach 2328-2336. [Crossref]

368. Guo-Jun Qi, Wei Liu, Charu Aggarwal, Thomas Huang. 2017. Joint Intermodal and Intramodal Label Transfers for Extremely Rare or Unseen Classes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**:7, 1360-1373. [Crossref]

369. Dimitrios Kollias, Mihalis A. Nicolaou, Irene Kotsia, Guoying Zhao, Stefanos Zafeiriou. Recognition of Affect in the Wild Using Deep Neural Networks 1972-1979. [Crossref]

370. Sadegh Mirshekarian, Razvan Bunescu, Cindy Marling, Frank Schwartz. Using LSTMs to learn physiological models of blood glucose behavior 2887-2891. [Crossref]

371. Chaochao Lu, Michael Hirsch, Bernhard Scholkopf. Flexible Spatio-Temporal Networks for Video Prediction 2137-2145. [Crossref]

372. Kuan-Lun Tseng, Yen-Liang Lin, Winston Hsu, Chung-Yang Huang. Joint Sequence Learning and Cross-Modality Convolution for 3D Biomedical Segmentation 3739-3746. [Crossref]

373. Oscar Koller, Sepehr Zargaran, Hermann Ney. Re-Sign: Re-Aligned End-to-End Sequence Modelling with Deep Recurrent CNN-HMMs 3416-3424. [Crossref]

374. Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Li Fei-Fei, C. Lawrence Zitnick, Ross Girshick. CLEVR: A Diagnostic Dataset for Compositional Language and Elementary Visual Reasoning 1988-1997. [Crossref]

375. Behrooz Mahasseni, Michael Lam, Sinisa Todorovic. Unsupervised Video Summarization with Adversarial LSTM Networks 2982-2991. [Crossref]

376. Kai Kang, Hongsheng Li, Tong Xiao, Wanli Ouyang, Junjie Yan, Xihui Liu, Xiaogang Wang. Object Detection in Videos with Tubelet Proposal Networks 889-897. [Crossref]

377. Zizhao Zhang, Yuanpu Xie, Fuyong Xing, Mason McGough, Lin Yang. MDNet: A Semantically and Visually Interpretable Medical Image Diagnosis Network 3549-3557. [Crossref]

378. Jonathan Krause, Justin Johnson, Ranjay Krishna, Li Fei-Fei. A Hierarchical Approach for Generating Descriptive Image Paragraphs 3337-3345. [Crossref]

379. Lu Huang, Ji Xu, Jiasong Sun, Yi Yang. An improved residual LSTM architecture for acoustic modeling 101-105. [Crossref]

380. Yufan Liu, Songyang Zhang, Mai Xu, Xuming He. Predicting Salient Face in Multiple-Face Videos 3224-3232. [Crossref]

381. Lorenzo Baraldi, Costantino Grana, Rita Cucchiara. Hierarchical Boundary-Aware Neural Encoder for Video Captioning 3185-3194. [Crossref]

382. Ronald Clark, Sen Wang, Andrew Markham, Niki Trigoni, Hongkai Wen. VidLoc: A Deep Spatio-Temporal Model for 6-DoF Video-Clip Relocalization 2652-2660. [Crossref]

383. Li Zhang, Tao Xiang, Shaogang Gong. Learning a Deep Embedding Model for Zero-Shot Learning 3010-3019. [Crossref]

384. Ruotian Luo, Gregory Shakhnarovich. Comprehension-Guided Referring Expressions 3125-3134. [Crossref]

385. Xuejian Rong, Chucai Yi, Yingli Tian. Unambiguous Text Localization and Retrieval for Cluttered Scenes 3279-3287. [Crossref]

386. Lei Zhang, Longxuan Ma. Coattention based BiLSTM for answer selection 1005-1011. [Crossref]

387. Linjie Yang, Kevin Tang, Jianchao Yang, Li-Jia Li. Dense Captioning with Joint Inference and Visual Context 1978-1987. [Crossref]

388. Shiliang Sun, Chen Luo, Junyu Chen. 2017. A review of natural language processing techniques for opinion mining systems. *Information Fusion* **36**, 10-25. [Crossref]

389. Yu Guan, Thomas Pl?tz. 2017. Ensembles of Deep LSTM Learners for Activity Recognition using Wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **1**:2, 1-28. [Crossref]

390. Ibrar Ahmad, Xiaojie Wang, Yuz hao Mao, Guang Liu, Haseeb Ahmad, Rahat Ullah. 2017. Ligature based Urdu Nastaleeq sentence recognition using gated bidirectional long short term memory. *Cluster Computing* **8**. . [Crossref]

391. Minlie Huang, Qiao Qian, Xiaoyan Zhu. 2017. Encoding Syntactic Knowledge in Neural Networks for Sentiment Classification. *ACM Transactions on Information Systems* **35**:3, 1-27. [Crossref]

392. Christos Melidis, Hiroyuki Iizuka, Davide Marocco. 2017. Intuitive control of mobile robots: an architecture for autonomous adaptive dynamic behaviour integration. *Cognitive Processing* **22**. . [Crossref]

393. Mohammad Nauman, Tamleek Ali Tanveer, Sohail Khan, Toqeer Ali Syed. 2017. Deep neural architectures for large scale android malware analysis. *Cluster Computing* **9**. . [Crossref]

394. Daniel Durstewitz. 2017. A state space approach for piecewise-linear recurrent neural networks for identifying computational dynamics from neural measurements. *PLOS Computational Biology* **13**:6, e1005542. [Crossref]

395. Miguel Ballesteros, Chris Dyer, Yoav Goldberg, Noah A. Smith. 2017. Greedy Transition-Based Dependency Parsing with Stack LSTMs. *Computational Linguistics* **43**:2, 311-347. [Abstract] [Full Text] [PDF] [PDF Plus]

396. Ning Liu, Zhe Li, Jielong Xu, Zhiyuan Xu, Sheng Lin, Qinru Qiu, Jian Tang, Yanzhi Wang. A Hierarchical Framework of Cloud Resource Allocation and Power Management Using Deep Reinforcement Learning 372-382. [Crossref]

397. Johannes Ziegmann, Jieqing Shi, Tobias Schnorer, Christian Endisch. Analysis of individual driver velocity prediction using data-driven driver models with environmental features 517-522. [Crossref]

398. , , , . 2017. LSTM-CRF for Drug-Named Entity Recognition. *Entropy* **19**:6, 283. [Crossref]

399. Maher Sameen, Biswajeet Pradhan. 2017. Severity Prediction of Traffic Accidents with Recurrent Neural Networks. *Applied Sciences* **7**:6, 476. [Crossref]

400. Luyang Li, Bing Qin, Ting Liu. 2017. Contradiction Detection with Contradiction-Specific Word Embedding. *Algorithms* **10**:2, 59. [Crossref]

401. Quan Nguyen Hong, Nghia Nguyen Tuan, Trung Tran Quang, Dung Nguyen Tien, Cuong Vo Le. Deep spatio-temporal network for accurate person re-identification 208-213. [Crossref]

402. Jitong Chen, DeLiang Wang. 2017. Long short-term memory for speaker generalization in supervised speech separation. *The Journal of the Acoustical Society of America* **141**:6, 4705-4714. [Crossref]

403. Saeeda Naz, Arif I. Umar, Riaz Ahmad, Imran Siddiqi, Saad B Ahmed, Muhammad I. Razzak, Faisal Shafait. 2017. Urdu Nastaliq recognition using convolutional–recursive deep learning. *Neurocomputing* **243**, 80-87. [Crossref]

404. David B. Brough, Abhiram Kannan, Benjamin Haaland, David G. Bucknall, Surya R. Kalidindi. 2017. Extraction of Process-Structure Evolution Linkages from X-ray Scattering Measurements Using Dimensionality Reduction and Time Series Analysis. *Integrating Materials and Manufacturing Innovation* **6**:2, 147-159. [Crossref]

405. Fan Wu, Kun Fu, Yang Wang, Zhibin Xiao, Xingyu Fu. 2017. A Spatial-Temporal-Semantic Neural Network Algorithm for Location Prediction on Moving Objects. *Algorithms* **10**:2, 37. [Crossref]

406. Yuhai Yu, Hongfei Lin, Jiana Meng, Xiaocong Wei, Zhehuan Zhao. 2017. Assembling Deep Neural Networks for Medical Compound Figure Detection. *Information* **8**:2, 48. [Crossref]

407. Emre Cakir, Giambattista Parascandolo, Toni Heittola, Heikki Huttunen, Tuomas Virtanen. 2017. Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:6, 1291-1303. [Crossref]

408. Guido Borghi, Riccardo Gasparini, Roberto Vezzani, Rita Cucchiara. Embedded recurrent network for head pose estimation in car 1503-1508. [Crossref]

409. V. John, S. Mita, H. Tehrani, K. Ishimaru. Automated driving by monocular camera using deep mixture of experts 127-134. [Crossref]

410. Yuya Jeremy Ong, Mu Qiao, Ramani Routray, Roger Raphael. Context-Aware Data Loss Prevention for Cloud Storage Services 399-406. [Crossref]

411. Pierre Nagorny, Maurice Pillet, Eric Pairel, Ronan Le Goff, Jerome Loureaux, Marlene Wali, Patrice Kiener. Quality prediction in injection molding 141-146. [Crossref]

412. Yumeng Wang, Wujun Che, Bo Xu. 2017. Encoder–decoder recurrent network model for interactive character animation generation. *The Visual Computer* **33**:6-8, 971-980. [Crossref]

413. Alex Zyner, Stewart Worrall, James Ward, Eduardo Nebot. Long short term memory for driver intent prediction 1484-1489. [Crossref]

414. Derek J. Phillips, Tim A. Wheeler, Mykel J. Kochenderfer. Generalizable intention prediction of human drivers at intersections 1665-1670. [Crossref]

415. Yona Falinie A. Gaus, Hongying Meng, Asim Jan. Decoupling Temporal Dynamics for Naturalistic Affect Recognition in a Two-Stage Regression Framework 1-6. [Crossref]

416. Keith Campbell, Wei Zuo, Deming Chen. 2017. New advances of high-level synthesis for efficient and reliable hardware design. *Integration, the VLSI Journal* **58**, 189-214. [Crossref]

417. Nicholas G. Polson, Vadim O. Sokolov. 2017. Deep learning for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies* **79**, 1-17. [Crossref]

418. Wenli Yu, Li Li, Xiaofei Xu, Dengbao Wang, Jingyuan Wang, Shiping Chen. ProductRec: Product Bundle Recommendation Based on User's Sequential Patterns in Social Networking Service Environment 301-308. [Crossref]

419. Jianhui Chen, James J. Little. 2017. Where should cameras look at soccer games: Improving smoothness using the overlapped hidden Markov model. *Computer Vision and Image Understanding* **159**, 59-73. [Crossref]

420. Thomas Wenzel, Steffen Brueggert, Joachim Denzler. Towards unconstrained content recognition of additional traffic signs 1421-1427. [Crossref]

421. Bo Zhao, Xiao Wu, Jiashi Feng, Qiang Peng, Shuicheng Yan. 2017. Diversified Visual Attention Networks for Fine-Grained Object Classification. *IEEE Transactions on Multimedia* **19**:6, 1245-1256. [Crossref]

422. Shuai Wang, Xiang Zhao, Bo Li, Bin Ge, Daquan Tang. Integrating Extractive and Abstractive Models for Long Text Summarization 305-312. [Crossref]

423. Cheng Feng, Tingting Li, Deeph Chana. Multi-level Anomaly Detection in Industrial Control Systems via Package Signatures and LSTM Networks 261-272. [Crossref]

424. Venice Erin Liong, Jiwen Lu, Yap-Peng Tan, Jie Zhou. 2017. Deep Video Hashing. *IEEE Transactions on Multimedia* **19**:6, 1209-1219. [Crossref]

425. Carlos Eduardo Cancino-Chacón, Thassilo Gadermaier, Gerhard Widmer, Maarten Grachten. 2017. An evaluation of linear and non-linear models of expressive dynamics in classical piano and symphonic music. *Machine Learning* **106**:6, 887-909. [Crossref]

426. Fei Li, Meishan Zhang, Bo Tian, Bo Chen, Guohong Fu, Donghong Ji. 2017. Recognizing irregular entities in biomedical text via deep neural networks. *Pattern Recognition Letters* . [Crossref]

427. Hongzhi Zhang, Weili Zhang, Tinglei Huang, Xiao Liang, Kun Fu. 2017. A Two-Stage Joint Model for Domain-Specific Entity Detection and Linking Leveraging an Unlabeled Corpus. *Information* **8**:2, 59. [Crossref]

428. Mourad Gridach. 2017. Character-level neural network for biomedical named entity recognition. *Journal of Biomedical Informatics* **70**, 85-91. [Crossref]

429. Xiao Yang, Dafang He, Wenyi Huang, Alexander Ororbia, Zihan Zhou, Daniel Kifer, C. Lee Giles. Smart Library: Identifying Books on Library Shelves Using Supervised Deep Learning for Scene Text Reading 1-4. [Crossref]

430. , , , , , . 2017. Identifying High Quality Document?Summary Pairs through Text Matching. *Information* **8**:2, 64. [Crossref]

431. Xianghua Fu, Wangwang Liu, Yingying Xu, Laizhong Cui. 2017. Combine HowNet lexicon to train phrase recursive autoencoder for sentence-level sentiment analysis. *Neurocomputing* **241**, 18-27. [Crossref]

432. Richard Gruner, Philip Henzler, Gereon Hinz, Corinna Eckstein, Alois Knoll. Spatiotemporal representation of driving scenarios and classification using neural networks 1782-1788. [Crossref]

433. Shuai Zheng, Kosta Ristovski, Ahmed Farahat, Chetan Gupta. Long Short-Term Memory Network for Remaining Useful Life estimation 88-95. [Crossref]

434. Tariqul Islam, Dakshnamoorthy Manivannan. Predicting Application Failure in Cloud: A Machine Learning Approach 24-31. [Crossref]

435. Pablo Arnau-González, Miguel Arevalillo-Herráez, Naeem Ramzan. 2017. Fusing highly dimensional energy and connectivity features to identify affective states from EEG signals. *Neurocomputing* **244**, 81-89. [Crossref]

436. Fei-yan Zhou, Lin-peng Jin, Jun Dong. 2017. Premature ventricular contraction detection combining deep neural networks and rules inference. *Artificial Intelligence in Medicine* . [Crossref]

437. Tomohiro Fujita, Wenjun Bai, Changqin Quan. Long short-term memory networks for automatic generation of conversations 483-487. [Crossref]

438. Alan Ferrari, Vanni Galli, Daniele Puccinelli, Silvia Giordano. On the usage of smart devices to augment the user interaction with multimedia applications 1-9. [Crossref]

439. Reza Mahjourian, Martin Wicke, Anelia Angelova. Geometry-based next frame prediction from monocular video 1700-1707. [Crossref]

440. Hao Chen, Lingyun Wu, Qi Dou, Jing Qin, Shengli Li, Jie-Zhi Cheng, Dong Ni, Pheng-Ann Heng. 2017. Ultrasound Standard Plane Detection Using a Composite Neural Network Framework. *IEEE Transactions on Cybernetics* **47**:6, 1576-1586. [Crossref]

441. Rong Zhang, Weiping Li, Wei Tan, Tong Mo. Deep and Shallow Model for Insurance Churn Prediction Service 346-353. [Crossref]

442. Arturo Gomez Chavez, Christian A. Mueller, Andreas Birk, Anja Babic, Nikola Miskovic. Stereo-vision based diver pose estimation using LSTM recurrent neural networks for AUV navigation guidance 1-7. [Crossref]

443. Karan K. Budhraja, John Winder, Tim Oates. 2017. Feature Construction for Controlling Swarms by Visual Demonstration. *ACM Transactions on Autonomous and Adaptive Systems* **12**:2, 1-22. [Crossref]

444. Jingwei Ma, Guang Li, Mingyang Zhong, Xin Zhao, Lei Zhu, Xue Li. 2017. LGA: latent genre aware micro-video recommendation on social media. *Multimedia Tools and Applications* **9**. . [Crossref]

445. Serena Yeung, Olga Russakovsky, Ning Jin, Mykhaylo Andriluka, Greg Mori, Li Fei-Fei. 2017. Every Moment Counts: Dense Detailed Labeling of Actions in Complex Videos. *International Journal of Computer Vision* **9**. . [Crossref]

446. Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, Joel T. Dudley. 2017. Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics* . [Crossref]

447. Bing Zhang, Jhen-Long Wu, Pei-Chann Chang. 2017. A multiple time series-based recurrent neural network for short-term load forecasting. *Soft Computing* **8**. . [Crossref]

448. Simone Scardapane, Lucas Stoffl, Florian Rohrbein, Aurelio Uncini. On the use of deep recurrent neural networks for detecting audio spoofing attacks 3483-3490. [Crossref]

449. Naifan Zhuang, Tuoerhongjiang Yusufu, Jun Ye, Kien A. Hua. Group Activity Recognition with Differential Recurrent Convolutional Neural Networks 526-531. [Crossref]

450. Robin Nix, Jian Zhang. Classification of Android apps and malware using deep neural networks 1871-1878. [Crossref]

451. Noriaki Hirose, Ryosuke Tajima. Modeling of rolling friction by recurrent neural network using LSTM 6471-6478. [Crossref]

452. Emanuele Principi, Fabio Vesperini, Stefano Squartini, Francesco Piazza. Acoustic novelty detection with adversarial autoencoders 3324-3330. [Crossref]

453. Ruiqi Hu, Celina Ping Yu, Sai-Fu Fung, Shirui Pan, Haishuai Wang, Guodong Long. Universal network representation for heterogeneous information networks 388-395. [Crossref]

454. Andros Tjandra, Sakriani Sakti, Satoshi Nakamura. Compressing recurrent neural network with tensor train 4451-4458. [Crossref]

455. Zhenghao Chen, Jianlong Zhou, Xiuying Wang, Jeremy Swanson, Fang Chen, Dagan Feng. Neural net-based and safety-oriented visual analytics for time-spatial data 1133-1140. [Crossref]

456. Trang Pham, Truyen Tran, Dinh Phung, Svetha Venkatesh. 2017. Predicting healthcare trajectories from medical records: A deep learning approach. *Journal of Biomedical Informatics* **69**, 218-229. [Crossref]

457. Yong Xu, Qiuqiang Kong, Qiang Huang, Wenwu Wang, Mark D. Plumbley. Convolutional gated recurrent neural network incorporating spatial features for audio tagging 3461-3466. [Crossref]

458. Naoyuki Kanda, Xugang Lu, Hisashi Kawai. 2017. Maximum-*a-Posteriori*-Based Decoding for End-to-End Acoustic Models. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:5, 1023-1034. [Crossref]

459. David M. Q. Nelson, Adriano C. M. Pereira, Renato A. de Oliveira. Stock market's price movement prediction with LSTM neural networks 1419-1426. [Crossref]

460. Caglar Gulcehre, Jose Sotelo, Marcin Moczulski, Yoshua Bengio. A robust adaptive stochastic gradient method for deep learning 125-132. [Crossref]

461. Mohammadhani Fouladgar, Mostafa Parchami, Ramez Elmasri, Amir Ghaderi. Scalable deep traffic flow neural networks for urban traffic congestion prediction 2251-2258. [Crossref]

462. Jason M. Allred, Kaushik Roy. Convolving over time via recurrent connections for sequential weight sharing in neural networks 4444-4450. [Crossref]

463. Juarez Monteiro, Joao Paulo Aires, Roger Granada, Rodrigo C. Barros, Felipe Meneguzzi. Virtual guide dog: An application to support visually-impaired people through deep convolutional neural networks 2267-2274. [Crossref]

464. Hongzhen Wang, Ying Wang, Qian Zhang, Shiming Xiang, Chunhong Pan. 2017. Gated Convolutional Neural Network for Semantic Segmentation in High-Resolution Images. *Remote Sensing* **9**:5, 446. [Crossref]

465. Xingwei He, Hua Xu, Xiaomin Sun, Junhui Deng, Jia Li. ABiRCNN with neural tensor network for answer selection 2582-2589. [Crossref]

466. Behzad Hasani, Mohammad H. Mahoor. Spatio-Temporal Facial Expression Recognition Using Convolutional Neural Networks and Conditional Random Fields 790-795. [Crossref]

467. Daesik Kim, Myunggi Lee, Nojun Kwak. Matching video net: Memory-based embedding for video action recognition 432-438. [Crossref]

468. A. R. Varkonyi-Koczy, Balazs Tusor, J. T. Toth. Robust variable length data classification with extended sequential fuzzy indexing tables 1-6. [Crossref]

469. Andrew Pulver, Siwei Lyu. LSTM with working memory 845-851. [Crossref]

470. Zhen Xu, Bingquan Liu, Baoxun Wang, Chengjie Sun, Xiaolong Wang. Incorporating loose-structured knowledge into conversation modeling via recall-gate LSTM 3506-3513. [Crossref]

471. Khuong N. Nguyen, Yoonsuck Choe. Dynamic control using feedforward networks with adaptive delay and facilitating neural dynamics 2987-2994. [Crossref]

472. Moirangthem Dennis Singh, Minho Lee. Temporal hierarchies in multilayer gated recurrent neural networks for language models 2152-2157. [Crossref]

473. Maryam Asadi-Aghbolaghi, Albert Clapes, Marco Bellantonio, Hugo Jair Escalante, Victor Ponce-Lopez, Xavier Baro, Isabelle Guyon, Shohreh Kasaei, Sergio Escalera. A Survey on Deep Learning Based Approaches for Action and Gesture Recognition in Image Sequences 476-483. [Crossref]

474. Liang Guo, Naipeng Li, Feng Jia, Yaguo Lei, Jing Lin. 2017. A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing* **240**, 98-109. [Crossref]

475. Jeremy Morton, Tim A. Wheeler, Mykel J. Kochenderfer. 2017. Analysis of Recurrent Neural Networks for Probabilistic Modeling of Driver Behavior. *IEEE Transactions on Intelligent Transportation Systems* **18**:5, 1289-1298. [Crossref]

476. Mohaddeseh Bastan, Shahram Khadivi, Mohammad Mehdi Homayounpour. Neural Machine Translation on scarce-resource condition: A case-study on Persian-English 1485-1490. [Crossref]

477. Junpei Zhong, Angelo Cangelosi, Tetsuya Ogata. Toward abstraction from multi-modal data: Empirical studies on multiple time-scale recurrent models 3625-3632. [Crossref]

478. Wenzhen Yuan, Chenzhuo Zhu, Andrew Owens, Mandayam A. Srinivasan, Edward H. Adelson. Shape-independent hardness estimation using deep learning and a GelSight tactile sensor 951-958. [Crossref]

479. Yuchen Qiao, Kazuma Hashimoto, Akiko Eriguchi, Haixia Wang, Dongsheng Wang, Yoshimasa Tsuruoka, Kenjiro Taura. Cache Friendly Parallelization of Neural Encoder-Decoder Models Without Padding on Multi-core Architecture 437-440. [Crossref]

480. Monit Shah Singh, Vinaychandran Pondenkandath, Bo Zhou, Paul Lukowicz, Marcus Liwickit. Transforming sensor data to the image domain for deep learning — An application to footstep detection 2665-2672. [Crossref]

481. Weijiang Feng, Naiyang Guan, Yuan Li, Xiang Zhang, Zhigang Luo. Audio visual speech recognition with multimodal recurrent neural networks 681-688. [Crossref]

482. Jonathan A. Cox. Parameter compression of recurrent neural networks and degradation of short-term memory 867-872. [Crossref]

483. Stefan Oehmcke, Oliver Zielinski, Oliver Kramer. Recurrent neural networks and exponential PAA for virtual marine sensors 4459-4466. [Crossref]

484. Joonatas Wehrmann, Willian Becker, Henry E. L. Cagnini, Rodrigo C. Barros. A character-based convolutional neural network for language-agnostic Twitter sentiment analysis 2384-2391. [Crossref]

485. Deger Ayata, Yusuf Yaslan, Mustafa Kamasak. Multi channel brain EEG signals based emotional arousal classification with unsupervised feature learning using autoencoders 1-4. [Crossref]

486. Hugh Rabagliati, Leonidas A.A. Doumas, Douglas K. Bemis. 2017. Representing composed meanings through temporal binding. *Cognition* **162**, 61-72. [Crossref]

487. Zackory Erickson, Alexander Clegg, Wenhao Yu, Greg Turk, C. Karen Liu, Charles C. Kemp. What does the person feel? Learning to infer applied forces during robot-assisted dressing 6058-6065. [Crossref]

488. Olaoluwa Adigun, Bart Kosko. Using noise to speed up video classification with recurrent backpropagation 108-115. [Crossref]

489. Jin Guo, Jinghui Cheng, Jane Cleland-Huang. Semantically Enhanced Software Traceability Using Deep Learning Techniques 3-14. [Crossref]

490. Tuan Anh Le, Atilim Giines Baydin, Robert Zinkov, Frank Wood. Using synthetic data to train neural networks is model-based reasoning 3514-3521. [Crossref]

491. Jun He, Dongliang Li, Bin Yang, Siming Cao, Bo Sun, Lejun Yu. Multi View Facial Action Unit Detection Based on CNN and BLSTM-RNN 848-853. [Crossref]

492. Zae Myung Kim, Hyungrai Oh, Han-Gyu Kim, Chae-Gyun Lim, Kyo-Joong Oh, Ho-Jin Choi. 2017. Modeling long-term human activeness using recurrent neural networks for biometric data. *BMC Medical Informatics and Decision Making* **17**:S1. . [Crossref]

493. Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A. Shamma, Michael S. Bernstein, Li Fei-Fei. 2017. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations. *International Journal of Computer Vision* **123**:1, 32-73. [Crossref]

494. Xinyu Fu, Eugene Ch'ng, Uwe Aickelin, Simon See. CRNN: A Joint Neural Network for Redundancy Detection 1-8. [Crossref]

495. Ryan Burt, Mihael Cudic, Jose C. Principe. Fusing attention with visual question answering 949-953. [Crossref]

496. Chunchun Yang, Zhiheng Li, Bin Liang, Weining Lu, Xueqian Wang, Houde Liu. A particle filter and long short term memory fusion algorithm for failure prognostic of proton exchange membrane fuel cells 5646-5651. [Crossref]

497. Esra Ergun, Filiz Gurkan, Onur Kaplan, Bilge Gunsel. Video action classification by deep learning 1-4. [Crossref]

498. Yu Wang. A new concept using LSTM Neural Networks for dynamic system identification 5324-5329. [Crossref]

499. Martin Bach-Andersen, Bo Rømer-Odgaard, Ole Winther. 2017. Flexible non-linear predictive models for large-scale wind turbine diagnostics. *Wind Energy* **20**:5, 753-764. [Crossref]

500. Filippo Maria Bianchi, Michael Kampffmeyer, Enrico Maiorino, Robert Jenssen. Temporal overdrive recurrent neural network 4275-4282. [Crossref]

501. B. Chandra, Rajesh Kumar Sharma. On improving recurrent neural network for image classification 1904-1907. [Crossref]

502. Juarez Monteiro, Roger Granada, Rodrigo C. Barros, Felipe Meneguzzi. Deep neural networks for kitchen activity recognition 2048-2055. [Crossref]

503. Oliver Wallscheid, Wilhelm Kirchgassner, Joachim Bocker. Investigation of long short-term memory networks to temperature prediction for permanent magnet synchronous motors 1940-1947. [Crossref]

504. Jiangjiang Liu, Biao Luo, Pengfei Yan, Ding Wang, Derong Liu. Long short-term memory based on a reward/punishment strategy for recurrent neural networks 327-332. [Crossref]

505. Atsunori Ogawa, Takaaki Hori. 2017. Error detection and accuracy estimation in automatic speech recognition using deep bidirectional recurrent neural networks. *Speech Communication* **89**, 70-83. [Crossref]

506. Marsela Polic, Ziad Salem, Karlo Griparic, Stjepan Bogdan, Thomas Schmickl. Estimation of moving agents density in 2D space based on LSTM neural network 1-8. [Crossref]

507. Hayk Shoukourian, Torsten Wilde, Detlef Labrenz, Arndt Bode. Using Machine Learning for Data Center Cooling Infrastructure Efficiency Prediction 954-963. [Crossref]

508. Nermin Samet, Samet Hicsonmez, Pinar Duygulu, Emre Akbas. Could we create a training set for image captioning using automatic translation? 1-4. [Crossref]

509. Shiyao Wang, Zhidong Deng. Tightly-coupled convolutional neural network with spatial-temporal memory for text classification 2370-2376. [Crossref]

510. Xiaoyong Yuan, Chuanhuang Li, Xiaolin Li. DeepDefense: Identifying DDoS Attack via Deep Learning 1-8. [Crossref]

511. Sadaf Gulshad, Jong-Hwan Kim. Deep convolutional and recurrent writer 2836-2842. [Crossref]

512. Xiaoqing Xu, Shouyi Yin, Peng Ouyang. Fast and low-power behavior analysis on vehicles using smartphones 1-4. [Crossref]

513. William B. Langdon, Shin Yoo, Mark Harman. Inferring Automatic Test Oracles 5-6. [Crossref]

514. Cristina Heghedus. PhD Forum: Forecasting Public Transit Using Neural Network Models 1-2. [Crossref]

515. Abdalraouf Hassan, Ausif Mahmood. Deep learning for sentence classification 1-5. [Crossref]

516. Xiang Jiang, Xuan Liu, Erico N de Souza, Baifan Hu, Daniel L. Silver, Stan Matwin. Improving point-based AIS trajectory classification with partition-wise gated recurrent units 4044-4051. [Crossref]

517. Rodrigo F. Berriel, Andre Teixeira Lopes, Alexandre Rodrigues, Flavio Miguel Varejao, Thiago Oliveira-Santos. Monthly energy consumption forecast: A deep learning approach 4283-4290. [Crossref]

518. Arthur Jack Russell, Emmanouil Benetos, Artur d'Avila Garcez. On the memory properties of recurrent neural models 2596-2603. [Crossref]

519. Weibo Gong, Yu-Chi Ho. On fast retrieval of relational experiences 3206-3211. [Crossref]

520. Taesik Na, Jong Hwan Ko, Jaeha Kung, Saibal Mukhopadhyay. On-chip training of recurrent neural networks with limited numerical precision 3716-3723. [Crossref]

521. Connor Schenck, Dieter Fox. Visual closed-loop control for pouring liquids 2629-2636. [Crossref]

522. Qichuan Yang, Zhiqiang He, Fujiang Ge, Yang Zhang. Sequence-to-sequence prediction of personal computer software by recurrent neural network 934-940. [Crossref]

523. Zejian Shi, Minyong Shi, Chunfang Li. The prediction of character based on recurrent neural network language model 613-616. [Crossref]

524. Younggi Kim, Jihoon An, Minseok Lee, Younghee Lee. An Activity-Embedding Approach for Next-Activity Prediction in a Multi-User Smart Space 1-6. [Crossref]

525. Chandranath Adak, Bidyut B. Chaudhuri, Michael Blumenstein. Impact of struck-out text on writer identification 1465-1471. [Crossref]

526. Yuting Wu, Mei Yuan, Shaopeng Dong, Li Lin, Yingqi Liu. 2017. Remaining Useful Life Estimation of Engineered Systems using vanilla LSTM Neural Networks. *Neurocomputing* . [Crossref]

527. Shuohao Li, Kota Yamaguchi. Attention to describe products with attributes 215-218. [Crossref]

528. S. Oprea, A. Garcia-Garcia, J. Garcia-Rodriguez, S. Orts-Escolano, M. Cazorla. A recurrent neural network based Schaeffer gesture recognition system 425-431. [Crossref]

529. Anna Rohrbach, Atousa Torabi, Marcus Rohrbach, Niket Tandon, Christopher Pal, Hugo Larochelle, Aaron Courville, Bernt Schiele. 2017. Movie Description. *International Journal of Computer Vision* **123**:1, 94-120. [Crossref]

530. Carol V. Alexandru, Sebastiano Panichella, Harald C. Gall. Replicating Parser Behavior Using Neural Machine Translation 316-319. [Crossref]

531. Dana Hughes, Alon Krauthammer, Nikolaus Correll. Recognizing social touch gestures using recurrent and convolutional neural networks 2315-2321. [Crossref]

532. Andre Xian Ming Chang, Eugenio Culurciello. Hardware accelerators for recurrent neural networks on FPGA 1-4. [Crossref]

533. Jing Wang, Jian Tang, Zhiyuan Xu, Yanzhi Wang, Guoliang Xue, Xing Zhang, Dejun Yang. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach 1-9. [Crossref]

534. Ahmad Arinaldi, Mohamad Ivan Fanany. Cheating video description based on sequences of gestures 1-6. [Crossref]

535. Binbin Song, Yao Yu, Yu Zhou, Ziqiang Wang, Sidan Du. 2017. Host load prediction with long short-term memory in cloud computing. *The Journal of Supercomputing* **5**. . [Crossref]

536. Biswajit Dev Sarma, S. R. Mahadeva Prasanna. 2017. Acoustic–Phonetic Analysis for Speech Recognition: A Review. *IETE Technical Review* **36**, 1-23. [Crossref]

537. Han Altae-Tran, Bharath Ramsundar, Aneesh S. Pappu, Vijay Pande. 2017. Low Data Drug Discovery with One-Shot Learning. *ACS Central Science* **3**:4, 283-293. [Crossref]

538. Xiaojia Pu, Gangshan Wu, Chunfeng Yuan. 2017. Exploring overall opinions for document level sentiment classification with structural SVM. *Multimedia Systems* **2**. . [Crossref]

539. Yoav Goldberg. 2017. Neural Network Methods for Natural Language Processing. *Synthesis Lectures on Human Language Technologies* **10**:1, 1-309. [Crossref]

540. Anastasia Ioannidou, Elisavet Chatzilari, Spiros Nikolopoulos, Ioannis Kompatsiaris. 2017. Deep Learning Advances in Computer Vision with 3D Data. *ACM Computing Surveys* **50**:2, 1-38. [Crossref]

541. Fu-Chen Chen, Mohammad R. Jahanshahi, Rih-Teng Wu, Chris Joffe. 2017. A texture-Based Video Processing Methodology Using Bayesian Data Fusion for Autonomous Crack Detection on Metallic Surfaces. *Computer-Aided Civil and Infrastructure Engineering* **32**:4, 271-287. [Crossref]

542. Yuan-ping Nie, Yi Han, Jiu-ming Huang, Bo Jiao, Ai-ping Li. 2017. Attention-based encoder-decoder model for answer selection in question answering. *Frontiers of Information Technology & Electronic Engineering* **18**:4, 535-544. [Crossref]

543. Samitha Herath, Mehrtash Harandi, Fatih Porikli. 2017. Going deeper into action recognition: A survey. *Image and Vision Computing* **60**, 4-21. [Crossref]

544. Xu Tao, Zhou Yun. 2017. Fall prediction based on biomechanics equilibrium using Kinect. *International Journal of Distributed Sensor Networks* **13**:4, 155014771770325. [Crossref]

545. Xu-Yao Zhang, Guo-Sen Xie, Cheng-Lin Liu, Yoshua Bengio. 2017. End-to-End Online Writer Identification With Recurrent Neural Network. *IEEE Transactions on Human-Machine Systems* **47**:2, 285-292. [Crossref]

546. Mengqi Pei, Xing Wu, Yike Guo, Hamido Fujita. 2017. Small bowel motility assessment based on fully convolutional networks and long short-term memory. *Knowledge-Based Systems* **121**, 163-172. [Crossref]

547. Jeff Donahue, Lisa Anne Hendricks, Marcus Rohrbach, Subhashini Venugopalan, Sergio Guadarrama, Kate Saenko, Trevor Darrell. 2017. Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**:4, 677-691. [Crossref]

548. Basura Fernando, Efstratios Gavves, Jose Oramas Oramas M., Amir Ghodrati, Tinne Tuytelaars. 2017. Rank Pooling for Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**:4, 773-787. [Crossref]

549. Tao Chen, Ruifeng Xu, Yulan He, Xuan Wang. 2017. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Systems with Applications* **72**, 221-230. [Crossref]

550. Ling Shao, Ziyun Cai, Li Liu, Ke Lu. 2017. Performance evaluation of deep feature learning for RGB-D image/video classification. *Information Sciences* **385-386**, 266-283. [Crossref]

551. Dilranjan S. Wickramasuriya, Calvin A. Perumalla, Kemal Davaslioglu, Richard D. Gitlin. Base station prediction and proactive mobility management in virtual cells using recurrent neural networks 1-6. [Crossref]

552. Shiliang Zhang, Cong Liu, Hui Jiang, Si Wei, Lirong Dai, Yu Hu. 2017. Nonrecurrent Neural Structure for Long-Term Dependence. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:4, 871-884. [Crossref]

553. Yafeng Ren, Donghong Ji. 2017. Neural networks for deceptive opinion spam detection: An empirical study. *Information Sciences* **385-386**, 213-224. [Crossref]

554. Jiaming Xu, Bo Xu, Peng Wang, Suncong Zheng, Guanhua Tian, Jun Zhao, Bo Xu. 2017. Self-Taught convolutional neural networks for short text clustering. *Neural Networks* **88**, 22-31. [Crossref]

555. Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan. 2017. Show and Tell: Lessons Learned from the 2015 MSCOCO Image Captioning Challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**:4, 652-663. [Crossref]

556. Yuguan Xing, Sosuke Amano, Toshihiko Yamasaki, Kiyoharu Aizawa, Makoto Ogawa. Distributed Representation of Dish Names in Food Related Web Services for Associative Search and Nutrition Estimation 82-85. [Crossref]

557. Andrej Karpathy, Li Fei-Fei. 2017. Deep Visual-Semantic Alignments for Generating Image Descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**:4, 664-676. [Crossref]

558. George Loukas, Yongpil Yoon, Georgia Sakellari, Tuan Vuong, Ryan Heartfield. 2017. Computation offloading of a vehicle's continuous intrusion detection workload for energy efficiency and performance. *Simulation Modelling Practice and Theory* **73**, 83-94. [Crossref]

559. Abdalraouf Hassan, Ausif Mahmood. Deep Learning approach for sentiment analysis of short texts 705-710. [Crossref]

560. Liren Chen, Ruijie Yan, Liangrui Peng, Akio Furuhata, Xiaoqing Ding. Multi-layer recurrent neural network based offline Arabic handwriting recognition 6-10. [Crossref]

561. Mohit Jain, Minesh Mathew, C. V. Jawahar. Unconstrained scene text and video text recognition for Arabic script 26-30. [Crossref]

562. Yan-Shuo Chang. 2017. Fine-grained attention for image caption generation. *Multimedia Tools and Applications* **27**. . [Crossref]

563. Georg Layher, Tobias Brosch, Heiko Neumann. 2017. Real-Time Biologically Inspired Action Recognition from Key Poses Using a Neuromorphic Architecture. *Frontiers in Neurorobotics* **11**. . [Crossref]

564. Ben Scheres, Wim H. van der Putten. 2017. The plant perceptron connects environment to development. *Nature* **543**:7645, 337-345. [Crossref]

565. Tom Charnock, Adam Moss. 2017. Deep Recurrent Neural Networks for Supernovae Classification. *The Astrophysical Journal* **837**:2, L28. [Crossref]

566. Jun Ye, Hao Hu, Guo-Jun Qi, Kien A. Hua. 2017. A Temporal Order Modeling Approach to Human Action Recognition from Multimodal Sensor Data. *ACM Transactions on Multimedia Computing, Communications, and Applications* **13**:2, 1-22. [Crossref]

567. Boyuan Deng, Denis Jouvet, Yves Laprie, Ingmar Steiner, Aghilas Sini. Towards confidence measures on fundamental frequency estimations 5605-5609. [Crossref]

568. Wei-Jen Ko, Bo-Hsiang Tseng, Hung-Yi Lee. Recurrent Neural Network based language modeling with controllable external Memory 5705-5709. [Crossref]

569. Shagan Sah, Sourabh Kulhare, Allison Gray, Subhashini Venugopalan, Emily Prud'Hommeaux, Raymond Ptucha. Semantic Text Summarization of Long Videos 989-997. [Crossref]

570. Florian Eyben, Matthias Unfried, Gerhard Hagerer, Bjorn Schuller. Automatic multi-lingual arousal detection from voice applied to real product testing applications 5155-5159. [Crossref]

571. Yu Zhang, William Chan, Navdeep Jaitly. Very deep convolutional networks for end-to-end speech recognition 4845-4849. [Crossref]

572. Xing Hao, Guigang Zhang. 2017. Deep learning. *Encyclopedia with Semantic Computing and Robotic Intelligence* **01**:01, 1630018. [Crossref]

573. Bing Liu, Ian Lane. Dialog context language modeling with recurrent neural networks 5715-5719. [Crossref]

574. Jonathan Chang, Stefan Scherer. Learning representations of emotional speech with deep convolutional generative adversarial networks 2746-2750. [Crossref]

575. Yuzhou Liu, DeLiang Wang. Time and frequency domain long short-term memory for noise robust pitch tracking 5600-5604. [Crossref]

576. Runnan Li, Zhiyong Wu, Xunying Liu, Helen Meng, Lianhong Cai. Multi-task learning of structured output layer bidirectional LSTMS for speech synthesis 5510-5514. [Crossref]

577. Jing Han, Zixing Zhang, Fabien Ringeval, Bjorn Schuller. Prediction-based learning for continuous emotion recognition in speech 5005-5009. [Crossref]

578. Yi Zhu, Shawn Newsam. Efficient Action Detection in Untrimmed Videos via Multi-task Learning 197-206. [Crossref]

579. Yin Xian, Yunchen Pu, Zhe Gan, Liang Lu, Andrew Thompson. Adaptive DCTNet for audio signal classification 3999-4003. [Crossref]

580. Anurag Kumar, Bhiksha Raj, Ndapandula Nakashole. Discovering sound concepts and acoustic relations in text 631-635. [Crossref]

581. Jeff Craley, Thomas S. Murray, Daniel R. Mendat, Andreas G. Andreou. Action recognition using micro-Doppler signatures and a recurrent neural network 1-5. [Crossref]

582. Syed Suleman Abbas Zaidi, Marium Mehboob Ali, Fakhra Aftab, Yusra Shahid, Muhammad Khurram. Name spotting over low signal-to-noise ratio (SNR) using Blind Source Separation and Connectionist Temporal Classification 320-325. [Crossref]

583. Qing Li, Zhaofan Qiu, Ting Yao, Tao Mei, Yong Rui, Jiebo Luo. 2017. Learning hierarchical video representation for action recognition. *International Journal of Multimedia Information Retrieval* **6**:1, 85-98. [Crossref]

584. Yong Zhao, Jinyu Li, Kshitiz Kumar, Yifan Gong. Extended low-rank plus diagonal adaptation for deep and recurrent neural networks 5040-5044. [Crossref]

585. Tim de Bruin, Kim Verbert, Robert Babuska. 2017. Railway Track Circuit Fault Diagnosis Using Recurrent Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems* **28**:3, 523-533. [Crossref]

586. Yuqing Wan, Yain-Whar Si. 2017. Adaptive neuro fuzzy inference system for chart pattern matching in financial time series. *Applied Soft Computing* . [Crossref]

587. Taro Ishitaki, Ryoichiro Obukata, Tetsuya Oda, Leonard Barolli. Application of Deep Recurrent Neural Networks for Prediction of User Behavior in Tor Networks 238-243. [Crossref]

588. W. Xiong, J. Droppo, X. Huang, F. Seide, M. Seltzer, A. Stolcke, D. Yu, G. Zweig. The microsoft 2016 conversational speech recognition system 5255-5259. [Crossref]

589. Tom Ko, Vijayaditya Peddinti, Daniel Povey, Michael L. Seltzer, Sanjeev Khudanpur. A study on data augmentation of reverberant speech for robust speech recognition 5220-5224. [Crossref]

590. Murali Karthick Baskar, Martin Karafiat, Lukas Burget, Karel Vesely, Frantisek Grezl, Jan Cernocky. Residual memory networks: Feed-forward approach to learn long-term temporal dependencies 4810-4814. [Crossref]

591. Kyuyeon Hwang, Wonyong Sung. Character-level language modeling with hierarchical recurrent neural networks 5720-5724. [Crossref]

592. Biplab Banerjee, Vittorio Murino. Efficient pooling of image based CNN features for action recognition in videos 2637-2641. [Crossref]

593. Enea Ceolini, Shih-Chii Liu. Impact of low-precision deep regression networks on single-channel source separation 256-260. [Crossref]

594. Nathan E West, Tim O'Shea. Deep architectures for modulation recognition 1-6. [Crossref]

595. Tom Sercu, George Saon, Jia Cui, Xiaodong Cui, Bhuvana Ramabhadran, Brian Kingsbury, Abhinav Sethy. Network architectures for multilingual speech representation learning 5295-5299. [Crossref]

596. Kun-Yi Huang, Chung-Hsien Wu, Ming-Hsiang Su, Hsiang-Chi Fu. Mood detection from daily conversational speech using denoising autoencoder and LSTM 5125-5129. [Crossref]

597. C. Stamate, G.D. Magoulas, S. Kueppers, E. Nomikou, I. Daskalopoulos, M.U. Luchini, T. Moussouri, G. Roussos. Deep learning Parkinson's from smartphone data 31-40. [Crossref]

598. Wenfu Wang, Bo Xu. Combining unidirectional long short-term memory with convolutional output layer for high-performance speech synthesis 5500-5504. [Crossref]

599. Jinyu Li, Yan Huang, Yifan Gong. Improved cepstra minimum-mean-square-error noise reduction algorithm for robust speech recognition 4865-4869. [Crossref]

600. Moustafa Alzantot, Supriyo Chakraborty, Mani Srivastava. SenseGen: A deep learning architecture for synthetic sensor data generation 188-193. [Crossref]

601. Asim Munawar, Phongtharin Vinayavekhin, Giovanni De Magistris. Spatio-Temporal Anomaly Detection for Industrial Robots through Prediction in Unsupervised Feature Space 1017-1025. [Crossref]

602. Xinhao Liu, Takahito Kawanishi, Xiaomeng Wu, Kaoru Hiramatsu, Kunio Kashino. Deep salience map guided arbitrary direction scene text recognition 1642-1646. [Crossref]

603. Artsiom Ablavatski, Shijian Lu, Jianfei Cai. Enriched Deep Recurrent Visual Attention Model for Multiple Object Recognition 971-978. [Crossref]

604. Yun Wang, Florian Metze. A first attempt at polyphonic sound event detection using connectionist temporal classification 2986-2990. [Crossref]

605. Albert Zeyer, Patrick Doetsch, Paul Voigtlaender, Ralf Schluter, Hermann Ney. A comprehensive study of deep bidirectional LSTM RNNS for acoustic modeling in speech recognition 2462-2466. [Crossref]

606. Jinhyeok Jang, Dae Hoe Kim, Hyung-Il Kim, Yong Man Ro. Color channel-wise recurrent learning for facial expression recognition 1233-1237. [Crossref]

607. Jian Zheng, Cencen Xu, Ziang Zhang, Xiaohua Li. Electric load forecasting in smart grids using Long-Short-Term-Memory based Recurrent Neural Network 1-6. [Crossref]

608. Harshala Gammulle, Simon Denman, Sridha Sridharan, Clinton Fookes. Two Stream LSTM: A Deep Fusion Framework for Human Action Recognition 177-186. [Crossref]

609. Ronald Salloum, C.-C. Jay Kuo. ECG-based biometrics using recurrent neural networks 2062-2066. [Crossref]

610. Patrick Doetsch, Albert Zeyer, Paul Voigtlaender, Ilia Kulikov, Ralf Schluter, Hermann Ney. Returnn: The RWTH extensible training framework for universal recurrent neural networks 5345-5349. [Crossref]

611. P. Nieters, J. Leugering, G. Pipa. 2017. Neuromorphic computation in multi-delay coupled models. *IBM Journal of Research and Development* **61**:2/3, 8:7-8:9. [Crossref]

612. Jia Cui, Brian Kingsbury, Bhuvana Ramabhadran, George Saon, Tom Sercu, Kartik Audhkhasi, Abhinav Sethy, Markus Nussbaum-Thom, Andrew Rosenberg. Knowledge distillation across ensembles of multilingual models for low-resource languages 4825-4829. [Crossref]

613. Ben Athiwaratkun, Jack W. Stokes. Malware classification with LSTM and GRU language models and a character-level CNN 2482-2486. [Crossref]

614. Xuesong Yang, Yun-Nung Chen, Dilek Hakkani-Tur, Paul Crook, Xiujun Li, Jianfeng Gao, Li Deng. End-to-end joint learning of natural language understanding and dialogue manager 5690-5694. [Crossref]

615. Juncheng Li, Wei Dai, Florian Metze, Shuhui Qu, Samarjit Das. A comparison of Deep Learning methods for environmental sound detection 126-130. [Crossref]

616. Geoffrey Zweig, Chengzhu Yu, Jasha Droppo, Andreas Stolcke. Advances in all-neural speech recognition 4805-4809. [Crossref]

617. Vladimir Rybalkin, Norbert Wehn, Mohammad Reza Yousefi, Didier Stricker. Hardware architecture of Bidirectional Long Short-Term Memory Neural Network for Optical Character Recognition 1390-1395. [Crossref]

618. Zhuo Chen, Yi Luo, Nima Mesgarani. Deep attractor network for single-microphone speaker separation 246-250. [Crossref]

619. Seonghan Ryu, Seokhwan Kim, Junhwi Choi, Hwanjo Yu, Gary Geunbae Lee. 2017. Neural sentence embedding using only in-domain sentences for out-of-domain sentence detection in dialog systems. *Pattern Recognition Letters* **88**, 26-32. [Crossref]

620. Yuki Saito, Shinnosuke Takamichi, Hiroshi Saruwatari. Training algorithm to deceive Anti-Spoofing Verification for DNN-based speech synthesis 4900-4904. [Crossref]

621. Zhiqiang Lv, Jian Kang, Wei-Qiang Zhang, Jia Liu. An LSTM-CTC based verification system for proxy-word based OOV keyword search 5655-5659. [Crossref]

622. Preethi Jyothi, Mark Hasegawa-Johnson. Low-resource grapheme-to-phoneme conversion using recurrent neural networks 5030-5034. [Crossref]

623. Jun Li, Xue Mei, Danil Prokhorov, Dacheng Tao. 2017. Deep Neural Network for Structural Prediction and Lane Detection in Traffic Scene. *IEEE Transactions on Neural Networks and Learning Systems* **28**:3, 690-703. [Crossref]

624. Amit Das, Ivan Tashev, Shoaib Mohammed. Ultrasound based gesture recognition 406-410. [Crossref]

625. Zhong Meng, Shinji Watanabe, John R. Hershey, Hakan Erdogan. Deep long short-term memory adaptive beamforming networks for multichannel robust speech recognition 271-275. [Crossref]

626. Zhiyuan Tang, Ying Shi, Dong Wang, Yang Feng, Shiyue Zhang. Memory visualization for gated recurrent neural networks in speech recognition 2736-2740. [Crossref]

627. Shuo Hong Wang, Jingwen Zhao, Xiang Liu, Zhi-Ming Qian, Ye Liu, Yan Qiu Chen. 3D tracking swimming fish school with learned kinematic model using LSTM network 1068-1072. [Crossref]

628. Kazuki Irie, Pavel Golik, Ralf Schluter, Hermann Ney. Investigations on byte-level convolutional neural networks for language modeling in low resource speech recognition 5740-5744. [Crossref]

629. Shengdong Zhang, Soheil Bahrampour, Naveen Ramakrishnan, Lukas Schott, Mohak Shah. Deep learning on symbolic representations for large-scale heterogeneous time-series event prediction 5970-5974. [Crossref]

630. Songyang Zhang, Xiaoming Liu, Jun Xiao. On Geometric Features for Skeleton-Based Action Recognition Using Multilayer LSTM Networks 148-157. [Crossref]

631. Bolan Su, Shijian Lu. 2017. Accurate recognition of words in scenes without character segmentation using recurrent neural network. *Pattern Recognition* **63**, 397-405. [Crossref]

632. Zheng Zhao, Weihai Chen, Xingming Wu, Peter C. Y. Chen, Jingmeng Liu. 2017. LSTM network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems* **11**:2, 68-75. [Crossref]

633. Herve Bredin. TristouNet: Triplet loss for speaker turn embedding 5430-5434. [Crossref]

634. Zhe Gan, P. D. Singh, Ameet Joshi, Xiaodong He, Jianshu Chen, Jianfeng Gao, Li Deng. Character-level deep conflation for business data analytics 2222-2226. [Crossref]

635. Jing Han, Zixing Zhang, Fabien Ringeval, Bjorn Schuller. Reconstruction-error-based learning for continuous emotion recognition in speech 2367-2371. [Crossref]

636. Kanishka Rao, Hasim Sak. Multi-accent speech recognition with hierarchical grapheme based models 4815-4819. [Crossref]

637. Suyoun Kim, Takaaki Hori, Shinji Watanabe. Joint CTC-attention based end-to-end speech recognition using multi-task learning 4835-4839. [Crossref]

638. Daqian Wei, Bo Wang, Gang Lin, Dichen Liu, Zhaoyang Dong, Hesen Liu, Yilu Liu. 2017. Research on Unstructured Text Data Mining and Fault Classification Based on RNN-LSTM with Malfunction Inspection Report. *Energies* **10**:3, 406. [Crossref]

639. Sepehr Valipour, Mennatullah Siam, Martin Jagersand, Nilanjan Ray. Recurrent Fully Convolutional Networks for Video Segmentation 29-36. [Crossref]

640. Hao Wu, Saurabh Prasad. 2017. Convolutional Recurrent Neural Networks forHyperspectral Data Classification. *Remote Sensing* **9**:3, 298. [Crossref]

641. Ji Li, Yongfei Shen. Image describing based on bidirectional LSTM and improved sequence sampling 735-739. [Crossref]

642. Ji Li, Zhen Wang. Online car-hailing dispatch: Deep supply-demand gap forecast on spark 811-815. [Crossref]

643. Timothy Chappelly, Cristina Cifuentes, Padmanabhan Krishnan, Shlomo Gevay. Machine learning for finding bugs: An initial report 21-26. [Crossref]

644. Cornelia Fermüller, Fang Wang, Yezhou Yang, Konstantinos Zampogiannis, Yi Zhang, Francisco Barranco, Michael Pfeiffer. 2017. Prediction of Manipulation Actions. *International Journal of Computer Vision* . [Crossref]

645. Laurence T. Hunt, Benjamin Y. Hayden. 2017. A distributed, hierarchical and recurrent framework for reward-based choice. *Nature Reviews Neuroscience* **18**:3, 172-182. [Crossref]

646. Umut Güçlü, Marcel A. J. van Gerven. 2017. Modeling the Dynamics of Human Brain Activity with Recurrent Neural Networks. *Frontiers in Computational Neuroscience* **11**. . [Crossref]

647. Dae Hoe Kim, Seong Tae Kim, Jung Min Chang, Yong Man Ro. 2017. Latent feature representation with depth directional long-term recurrent learning for breast masses in digital breast tomosynthesis. *Physics in Medicine and Biology* **62**:3, 1009-1031. [Crossref]

648. Apoorva Bhandari, David Badre, Michael J. Frank. Learning Cognitive Control 376-391. [Crossref]

649. Tom Verguts. Computational Models of Cognitive Control 125-142. [Crossref]

650. Jonathan D. Cohen. Cognitive Control 1-28. [Crossref]

651. Rufin VanRullen. 2017. Perception Science in the Age of Deep Neural Networks. *Frontiers in Psychology* **8**. . [Crossref]

652. Saket Navlakha. 2017. Learning the Structural Vocabulary of a Network. *Neural Computation* **29**:2, 287-312. [Abstract] [Full Text] [PDF] [PDF Plus]

653. Atefeh Abdolmanafi, Luc Duong, Nagib Dahdah, Farida Cheriet. 2017. Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography. *Biomedical Optics Express* **8**:2, 1203. [Crossref]

654. Thi-Thu-Huong Le, Jihyun Kim, Howon Kim. An Effective Intrusion Detection Classifier Using Long Short-Term Memory with Gradient Descent Optimization 1-6. [Crossref]

655. Jung-Woo Ha, Adrian Kim, Dongwon Kim, Jeonghee Kim, Jeong-Whun Kim, Jin Joo Park, Borim Ryu. Predicting high-risk prognosis from diagnostic histories of adult disease patients via deep recurrent neural networks 394-399. [Crossref]

656. Marcel A.J. van Gerven. 2017. A primer on encoding models in sensory neuroscience. *Journal of Mathematical Psychology* **76**, 172-183. [Crossref]

657. Christoph Wick. 2017. Deep Learning. *Informatik-Spektrum* **40**:1, 103-107. [Crossref]

658. Ning An, Shi-Ying Sun, Xiao-Guang Zhao, Zeng-Guang Hou. 2017. Remember like humans. *International Journal of Advanced Robotic Systems* **14**:1, 172988141769231. [Crossref]

659. Han-Gyu Kim, Gil-Jin Jang, Ho-Jin Choi, Minho Kim, Young-Won Kim, Jaehun Choi. Recurrent neural networks with missing information imputation for medical examination data prediction 317-323. [Crossref]

660. Saeeda Naz, Arif I. Umar, Riaz Ahmad, Saad B. Ahmed, Syed H. Shirazi, Muhammad I. Razzak. 2017. Urdu Nasta'liq text recognition system based on multi-dimensional recurrent neural network and statistical features. *Neural Computing and Applications* **28**:2, 219-231. [Crossref]

661. Zifeng Wu, Yongzhen Huang, Liang Wang, Xiaogang Wang, Tieniu Tan. 2017. A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**:2, 209-226. [Crossref]

662. Shuohao Li, Jun Zhang, Qiang Guo, Jun Lei, Dan Tu. 2017. Generating image descriptions with multidirectional 2D long short-term memory. *IET Computer Vision* **11**:1, 104-111. [Crossref]

663. Diego Rueda Plata, Raúl Ramos-Pollán, Fabio A. González. 2017. Effective training of convolutional neural networks with small, specialized datasets. *Journal of Intelligent & Fuzzy Systems* **32**:2, 1333-1342. [Crossref]

664. Andreas Persson, Martin Längkvist, Amy Loutfi. 2017. Learning Actions to Improve the Perceptual Anchoring of Objects. *Frontiers in Robotics and AI* **3**. . [Crossref]

665. Marc Ebner. 2017. Distributed storage and recall of sentences. *Bio-Algorithms and Med-Systems* **13**:2. . [Crossref]

666. K. C. Santosh, Partha Pratim Roy. 2017. Arrow detection in biomedical images using sequential classifier. *International Journal of Machine Learning and Cybernetics* . [Crossref]

667. Derwin Suhartono, Aryo Pradipta Gema, Suhendro Winton, Theodorus David, Mohamad Ivan Fanany, Aniati Murni Arymurthy. Hierarchical Attention Network with XGBoost for Recognizing Insufficiently Supported Argument 174-188. [Crossref]

668. Yu Chen, Jian Yang, Jianjun Qian. 2017. Recurrent neural network for facial landmark detection. *Neurocomputing* **219**, 26-38. [Crossref]

669. Bin Kong, Xin Wang, Zhongyu Li, Qi Song, Shaoting Zhang. Cancer Metastasis Detection via Spatially Structured Deep Network 236-248. [Crossref]

670. Fan Yang, Jaymar Soriano, Takatomi Kubo, Kazushi Ikeda. A Hierarchical Mixture Density Network 878-885. [Crossref]

671. Yueting Zhuang, Hanqi Wang, Jun Xiao, Fei Wu, Yi Yang, Weiming Lu, Zhongfei Zhang. 2017. Bag-of-Discriminative-Words (BoDW) Representation via Topic Modeling. *IEEE Transactions on Knowledge and Data Engineering* 1-1. [Crossref]

672. Bander Alsulami, Edwin Dauber, Richard Harang, Spiros Mancoridis, Rachel Greenstadt. Source Code Authorship Attribution Using Long Short-Term Memory Based Networks 65-82. [Crossref]

673. Xue Yan, Wei-Long Zheng, Wei Liu, Bao-Liang Lu. Investigating Gender Differences of Brain Areas in Emotion Recognition Using LSTM Neural Network 820-829. [Crossref]

674. James J. Q. Yu, Albert Y. S. Lam, David J. Hill, Victor O. K. Li. 2017. Delay Aware Intelligent Transient Stability Assessment System. *IEEE Access* **5**, 17230-17239. [Crossref]

675. Yu Liu, Yanming Guo, Michael S. Lew. What Convnets Make for Image Captioning? 416-428. [Crossref]

676. Máté Ákos Tündik, Balázs Tarján, György Szaszák. Low Latency MaxEnt- and RNN-Based Word Sequence Models for Punctuation Restoration of Closed Caption Data 155-166. [Crossref]

677. Tomasz Trzci?ski, Pawe? Andruszkiewicz, Tomasz Boche?ski, Przemys?aw Rokita. Recurrent Neural Networks for Online Video Popularity Prediction 146-153. [Crossref]

678. Marcin Kuta, Mikołaj Morawiec, Jacek Kitowski. Sentiment Analysis with Tree-Structured Gated Recurrent Units 74-82. [Crossref]

679. Rory Finnegan, Mark Shaw, Suzanna Becker. Restricted Boltzmann Machine Models of Hippocampal Coding and Neurogenesis 443-461. [Crossref]

680. Mohammed Salah Al-Radhi, Tamás Gábor Csapó, Géza Németh. Deep Recurrent Neural Networks in Speech Synthesis Using a Continuous Vocoder 282-291. [Crossref]

681. Jarosław Protasiewicz, Marcin Mirończuk, Sławomir Dadas. Categorization of Multilingual Scientific Documents by a Compound Classification System 563-573. [Crossref]

682. Imen Grida Ben Yahia, Jaafar Bendriss, Alassane Samba, Philippe Dooze. CogNitive 5G networks: Comprehensive operator use cases with machine learning for management operations 252-259. [Crossref]

683. Younggi KIM, Younghee LEE. 2017. Next-Activity Set Prediction Based on Sequence Partitioning to Reduce Activity Pattern Complexity in the Multi-User Smart Space. *IEICE Transactions on Information and Systems* **E100.D**:10, 2587-2596. [Crossref]

684. Amir H. Abdi, Christina Luong, Teresa Tsang, John Jue, Ken Gin, Darwin Yeung, Dale Hawley, Robert Rohling, Purang Abolmaesumi. Quality Assessment of Echocardiographic Cine Using Recurrent Neural Networks: Feasibility on Five Standard View Planes 302-310. [Crossref]

685. Hainan Zhang, Yanyan Lan, Jiafeng Guo, Jun Xu, Xueqi Cheng. Neural or Statistical: An Empirical Study on Language Models for Chinese Input Recommendation on Mobile 3-16. [Crossref]

686. Dong Wu, Mingmin Chi. 2017. Long Short-Term Memory With Quadratic Connections in Recursive Neural Networks for Representing Compositional Semantics. *IEEE Access* **5**, 16077-16083. [Crossref]

687. Yishen Zhang, Dong Wang, Yuehui Chen, Huijie Shang, Qi Tian. Credit Risk Assessment Based on Long Short-Term Memory Model 700-712. [Crossref]

688. Maxim Tkachenko, Alexander Yamshinin, Nikolay Lyubimov, Mikhail Kotov, Marina Nastasenko. Speech Enhancement for Speaker Recognition Using Deep Recurrent Neural Networks 690-699. [Crossref]

689. Chuwei Luo, Wenjie Li, Qiang Chen, Yanxiang He. A Part-of-Speech Enhanced Neural Conversation Model 173-185. [Crossref]

690. Tomasz Kornuta, Kamil Rocki. Utilization of Deep Reinforcement Learning for Saccadic-Based Object Visual Search 565-574. [Crossref]

691. Alan Mosca, George D. Magoulas. Learning Input Features Representations in Deep Learning 433-445. [Crossref]

692. Hui Wen, Minglan Li, Zhili Ye. Neural Architecture for Negative Opinion Expressions Extraction 460-474. [Crossref]

693. Yao-zhong Zhang, Rui Yamaguchi, Seiya Imoto, Satoru Miyano. 2017. Sequence-specific bias correction for RNA-seq data using recurrent neural networks. *BMC Genomics* **18**:S1. . [Crossref]

694. Ning Xie, Kyle Brown, Nathan Rude, Derek Doran. A Soft Computing Prefetcher to Mitigate Cache Degradation by Web Robots 536-546. [Crossref]

695. Jian Hu, Xin Xin, Ping Guo. LSTM with Matrix Factorization for Road Speed Prediction 242-249. [Crossref]

696. Marco Bellantonio, Mohammad A. Haque, Pau Rodriguez, Kamal Nasrollahi, Taisi Telve, Sergio Escalera, Jordi Gonzalez, Thomas B. Moeslund, Pejman Rasti, Gholamreza Anbarjafari. Spatio-temporal Pain Recognition in CNN-Based Super-Resolved Facial Images 151-162. [Crossref]

697. Dmitry Devyatkin, Artem Shelmanov. Text Processing Framework for Emergency Event Detection in the Arctic Zone 74-88. [Crossref]

698. Jaafar Bendriss, Imen Grida Ben Yahia, Djamal Zeghlache. Forecasting and anticipating SLO breaches in programmable networks 127-134. [Crossref]

699. Connor Schenck, Dieter Fox. Towards Learning to Perceive and Reason About Liquids 488-501. [Crossref]

700. Natalia Viani, Timothy A. Miller, Dmitriy Dligach, Steven Bethard, Carlo Napolitano, Silvia G. Priori, Riccardo Bellazzi, Lucia Sacchi, Guergana K. Savova. Recurrent Neural Network Architectures for Event Extraction from Italian Medical Reports 198-202. [Crossref]

701. Weixuan Chen, Akane Sano, Daniel Lopez Martinez, Sara Taylor, Andrew W. McHill, Andrew J. K. Phillips, Laura Barger, Elizabeth B. Klerman, Rosalind W. Picard. Multimodal ambulatory sleep detection 465-468. [Crossref]

702. Makoto Kawano, Kazuhiro Ueda. 2017. Microblog Geolocation Estimation with Recurrent Neural Networks. *Transactions of the Japanese Society for Artificial Intelligence* **32**:1, WII-E_1-8. [Crossref]

703. Ronald M. Summers. Deep Learning and Computer-Aided Diagnosis for Medical Image Processing: A Personal Perspective 3-10. [Crossref]

704. Maryam Asadi-Aghbolaghi, Albert Clapés, Marco Bellantonio, Hugo Jair Escalante, Víctor Ponce-López, Xavier Baró, Isabelle Guyon, Shohreh Kasaei, Sergio Escalera. Deep Learning for Action and Gesture Recognition in Image Sequences: A Survey 539-578. [Crossref]

705. Fei Li, Meishan Zhang, Guohong Fu, Donghong Ji. A Neural Joint Model for Extracting Bacteria and Their Locations 15-26. [Crossref]

706. Ammar Mahmood, Mohammed Bennamoun, Senjian An, Ferdous Sohel, Farid Boussaid, Renae Hovey, Gary Kendrick, Robert B. Fisher. Deep Learning for Coral Classification 383-401. [Crossref]

707. Vincent Andrearczyk, Paul F. Whelan. Deep Learning in Texture Analysis and Its Application to Tissue Image Classification 95-129. [Crossref]

708. Kiran Ramesh, Surya Ravishankaran, Abhishek Joshi, K. Chandrasekaran. A Survey of Design Techniques for Conversational Agents 336-350. [Crossref]

709. Chao Wang, Ying Zhang, Wei Jie, Christian Sauer, Xiaojie Yuan. Hierarchical Semantic Representations of Online News Comments for Emotion Tagging Using Multiple Information Sources 121-136. [Crossref]

710. Ankesh Anand, Tanmoy Chakraborty, Noseong Park. We Used Neural Networks to Detect Clickbaits: You Won't Believe What Happened Next! 541-547. [Crossref]

711. Hongjun ZHANG, Yuntian FENG, Wenning HAO, Gang CHEN, Dawei JIN. 2017. Relation Extraction with Deep Reinforcement Learning. *IEICE Transactions on Information and Systems* **E100.D**:8, 1893-1902. [Crossref]

712. Erik Marchi, Fabio Vesperini, Stefano Squartini, Björn Schuller. 2017. Deep Recurrent Neural Network-Based Autoencoders for Acoustic Novelty Detection. *Computational Intelligence and Neuroscience* **2017**, 1-14. [Crossref]

713. Dazhen Lin, Donglin Cao, Yanping Lv, Zheng Cai. 2017. GIF Video Sentiment Detection Using Semantic Sequence. *Mathematical Problems in Engineering* **2017**, 1-11. [Crossref]

714. Hongning Wang, Rui Li, Milad Shokouhi, Hang Li, Yi Chang. 2017. Search, Mining, and Their Applications on Mobile Devices. *ACM Transactions on Information Systems* **35**:4, 1. [Crossref]

715. Andrés Pérez, Angélica Quevedo, Juan C. Caicedo. Computing Arithmetic Operations on Sequences of Handwritten Digits 393-400. [Crossref]

716. Qianzi Shen, Zijian Wang, Yaoru Sun. Sentiment Analysis of Movie Reviews Based on CNN-BLSTM 164-171. [Crossref]

717. Boyi Liu, Jieren Cheng, Kuanqi Cai, Pengchao Shi, Xiangyan Tang. Singular Point Probability Improve LSTM Network Performance for Long-term Traffic Flow Prediction 328-340. [Crossref]

718. Sebastian Otte, Theresa Schmitt, Karl Friston, Martin V. Butz. Inferring Adaptive Goal-Directed Behavior Within Recurrent Neural Networks 227-235. [Crossref]

719. Chiara Di Francescomarino, Chiara Ghidini, Fabrizio Maria Maggi, Giulio Petrucci, Anton Yeshchenko. An Eye into the Future: Leveraging A-priori Knowledge in Predictive Business Process Monitoring 252-268. [Crossref]

720. Zhen Huang, Sabato Marco Siniscalchi, Chin-Hui Lee. 2017. Bayesian Unsupervised Batch and Online Speaker Adaptation of Activation Function Parameters in Deep Models for Automatic Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **25**:1, 60-71. [Crossref]

721. Fatemeh Taheri Dezaki, Neeraj Dhungel, Amir H. Abdi, Christina Luong, Teresa Tsang, John Jue, Ken Gin, Dale Hawley, Robert Rohling, Purang Abolmaesumi. Deep Residual Recurrent Neural Networks for Characterisation of Cardiac Cycle Phase from Echocardiograms 100-108. [Crossref]

722. Minlong Lu, Zhiwei Deng, Ze-Nian Li. Learning Contextual Dependencies for Optical Flow with Recurrent Neural Networks 68-83. [Crossref]

723. Ashish Dandekar, Remmy A. M. Zen, Stéphane Bressan. Generating Fake but Realistic Headlines Using Deep Neural Networks 427-440. [Crossref]

724. YuKang Jia, Zhicheng Wu, Yanyan Xu, Dengfeng Ke, Kaile Su. 2017. Long Short-Term Memory Projection Recurrent Neural Network Architectures for Piano's Continuous Note Recognition. *Journal of Robotics* **2017**, 1. [Crossref]

725. Yu-xin Peng, Wen-wu Zhu, Yao Zhao, Chang-sheng Xu, Qing-ming Huang, Han-qing Lu, Qing-hua Zheng, Tie-jun Huang, Wen Gao. 2017. Cross-media analysis and reasoning: advances and directions. *Frontiers of Information Technology & Electronic Engineering* **18**:1, 44-57. [Crossref]

726. Umut Güçlü, Marcel A.J. van Gerven. 2017. Increasingly complex representations of natural movies across the dorsal stream are shared between subjects. *NeuroImage* **145**, 329-336. [Crossref]

727. Masashi Tsubaki, Masashi Shimbo, Yuji Matsumoto. 2017. Protein Fold Recognition with Representation Learning and Long Short-Term Memory. *IPSJ Transactions on Bioinformatics* **10**:0, 2-8. [Crossref]

728. Przemyslaw Buczkowski. Predicting Stock Trends Based on Expert Recommendations Using GRU/LSTM Neural Networks 708-717. [Crossref]

729. Tsung-Yen Yang, Christopher G. Brinton, Carlee Joe-Wong, Mung Chiang. 2017. Behavior-Based Grade Prediction for MOOCs via Time Series Neural Networks. *IEEE Journal of Selected Topics in Signal Processing* 1-1. [Crossref]

730. Md Mizanur Rahoman, Ryutaro Ichise. Relation-Wise Automatic Domain-Range Information Management for Knowledge Entries 105-108. [Crossref]

731. Xin Liu, Ryutaro Ichise. Food Sales Prediction with Meteorological Data ? A Case Study of a Japanese Chain Supermarket 93-104. [Crossref]

732. Damla Arifoglu, Abdelhamid Bouchachia. 2017. Activity Recognition and Abnormal Behaviour Detection with Recurrent Neural Networks. *Procedia Computer Science* **110**, 86-93. [Crossref]

733. Manasjyoti Bhuyan, Kandarpa Sarma, Nikos Mastorakis. 2017. Modified FLNN and Channel Sounder Arrangementbased Nonlinear Wireless Channel Modeling. *IEEE Access* 1-1. [Crossref]

734. Arman Cohan, Allan Fong, Nazli Goharian, Raj Ratwani. A Neural Attention Model for Categorizing Patient Safety Events 720-726. [Crossref]

735. Zhipeng Xie, Junfeng Hu. Max-Cosine Matching Based Neural Models for Recognizing Textual Entailment 295-308. [Crossref]

736. Ying Hua Tan, Chee Seng Chan. phi-LSTM: A Phrase-Based Hierarchical LSTM Model for Image Captioning 101-117. [Crossref]

737. Takuya SUZUKI, Yasutomo MATSUOKA. 2017. BASICAL STUDY ON HYSTERETIC MODELING BY USING RECURRENT NEURAL NETWORK. *Journal of Structural and Construction Engineering (Transactions of AIJ)* **82**:734, 543-553. [Crossref]

738. Julia Hofmockel, Felix Richter, Eric Sax. Automatic Defect Detection by One-Class Classification on Raw Vehicle Sensor Data 378-384. [Crossref]

739. Mihai Zanfir, Elisabeta Marinoiu, Cristian Sminchisescu. Spatio-Temporal Attention Models for Grounded Video Captioning 104-119. [Crossref]

740. Daehyun Lee, Jongmin Lee, Kee-Eung Kim. Multi-view Automatic Lip-Reading Using Neural Network 290-302. [Crossref]

741. Xiang Xiang, Trac D. Tran. Recursively Measured Action Units 153-159. [Crossref]

742. Louisa Pragst, Stefan Ultes, Wolfgang Minker. Recurrent Neural Network Interaction Quality Estimation 381-393. [Crossref]

743. Hui-Kang Yi, Jiu-Ming Huang, Shu-Qiang Yang. 2017. A Chinese Named Entity Recognition System with Neural Networks. *ITM Web of Conferences* **12**, 04002. [Crossref]

744. Giosué Lo Bosco, Mattia Antonino Di Gangi. Deep Learning Architectures for DNA Sequence Classification 162-171. [Crossref]

745. Wojciech Stokowiec, Tomasz Trzci?ski, Krzysztof Wo?k, Krzysztof Marasek, Przemys?aw Rokita. Shallow Reading with Deep Learning: Predicting Popularity of Online Content Using only Its Title 136-145. [Crossref]

746. Daniel Gerber, Stefan Meier, Walter Kellermann. Efficient target activity detection based on recurrent neural networks 46-50. [Crossref]

747. Jakub Nowak, Ahmet Taspinar, Rafał Scherer. LSTM Recurrent Neural Networks for Short Text and Sentiment Classification 553-562. [Crossref]

748. Dmitrii Fedotov, Maxim Sidorov, Wolfgang Minker. Context-Awared Models in Time-Continuous Multidimensional Affect Recognition 59-66. [Crossref]

749. Hounaïda Moalla, Walid Elloumi, Adel M. Alimi. H-PSO-LSTM: Hybrid LSTM Trained by PSO for Online Handwriter Identification 41-50. [Crossref]

750. Sebastian Otte, Adrian Zwiener, Martin V. Butz. Inherently Constraint-Aware Control of Many-Joint Robot Arms with Inverse Recurrent Models 262-270. [Crossref]

751. Sabrina Amrouche, Nils Braun, Paolo Calafiura, Steven Farrell, Jochen Gemmler, Cécile Germain, Vladimir Vava Gligorov, Tobias Golling, Heather Gray, Isabelle Guyon, Mikhail Hushchyn, Vincenzo Innocente, Balázs Kégl, Sara Neuhaus, David Rousseau, Andreas Salzburger, Andrei Ustyuzhanin, Jean-Roch Vlimant, Christian Wessel, Yetkin Yilmaz. 2017. Track reconstruction at LHC as a collaborative data challenge use case with RAMP. *EPJ Web of Conferences* **150**, 00015. [Crossref]

752. Daniel D. Johnson. Generating Polyphonic Music Using Tied Parallel Networks 128-143. [Crossref]

753. Siqi Bao, Pei Wang, Albert C. S. Chung. 3D Randomized Connection Network with Graph-Based Inference 47-55. [Crossref]

754. Fei Hu, Xiaofei Xu, Jingyuan Wang, Zhanbo Yang, Li Li. Memory-Enhanced Latent Semantic Model: Short Text Understanding for Sentiment Analysis 393-407. [Crossref]

755. Siamak Barzegar, Andre Freitas, Siegfried Handschuh, Brian Davis. Composite Semantic Relation Classification 406-417. [Crossref]

756. Fu-Hsiang Chan, Yu-Ting Chen, Yu Xiang, Min Sun. Anticipating Accidents in Dashcam Videos 136-153. [Crossref]

757. Jiangping Huang, Shuxin Yao, Chen Lyu, Donghong Ji. Multi-Granularity Neural Sentence Model for Measuring Short Text Similarity 439-455. [Crossref]

758. Pengda Qin, Weiran Xu, Jun Guo. 2017. Providing Definitive Learning Direction for Relation Classification System. *Journal of Control Science and Engineering* **2017**, 1-10. [Crossref]

759. Laurynas Dovydaitis, Vytautas Rudžionis. Identifying Lithuanian Native Speakers Using Voice Recognition 79-84. [Crossref]

760. Da Pan, Mengqi Wang, Meishan Zhang, Guohong Fu. Extracting Chinese Explanatory Expressions with Discrete and Neural CRFs 1-12. [Crossref]

761. Marina Zimmermann, Mostafa Mehdipour Ghazi, Hazım Kemal Ekenel, Jean-Philippe Thiran. Visual Speech Recognition Using PCA Networks and LSTMs in a Tandem GMM-HMM System 264-276. [Crossref]

762. Hao Tang, Wei Liu, Wei-Long Zheng, Bao-Liang Lu. Multimodal Emotion Recognition Using Deep Neural Networks 811-819. [Crossref]

763. Tu Ngoc Nguyen, Cheng Li, Claudia Niederée. On Early-Stage Debunking Rumors on Twitter: Leveraging the Wisdom of Weak Learners 141-158. [Crossref]

764. Deng Cai, Hai Zhao. Pair-Aware Neural Sentence Modeling for Implicit Discourse Relation Classification 458-466. [Crossref]

765. Kosuke Ota, Keiichiro Shirai, Hidetoshi Miyao, Minoru Maruyama. Interactive Image Search System Based on Multimodal Analogy 591-596. [Crossref]

766. . References 573-600. [Crossref]

767. D. Bug, S. Schneider, A. Grote, E. Oswald, F. Feuerhake, J. Schüler, D. Merhof. Context-Based Normalization of Histological Stains Using Deep Convolutional Features 135-142. [Crossref]

768. Oya Celiktutan, Hatice Gunes. 2017. Automatic Prediction of Impressions in Time and across Varying Context: Personality, Attractiveness and Likeability. *IEEE Transactions on Affective Computing* **8**:1, 29-42. [Crossref]

769. Anup K. Kalia, Jin Xiao, Muhammed F. Bulut, Maja Vukovic, Nikos Anerousis. Cataloger: Catalog Recommendation Service for IT Change Requests 545-560. [Crossref]

770. Daniel Durstewitz. 199. [Crossref]

771. Xue-Ren Sun, Shao-He Lv, Xiao-Dong Wang, Dong Wang. 2017. Chinese Word Sense Disambiguation using a LSTM. *ITM Web of Conferences* **12**, 01027. [Crossref]

772. Philip Hake, Manuel Zapp, Peter Fettke, Peter Loos. Supporting Business Process Modeling Using RNNs for Label Classification 283-286. [Crossref]

773. Moritz August, Xiaotong Ni. 2017. Using recurrent neural networks to optimize dynamical decoupling for quantum memory. *Physical Review A* **95**:1. . [Crossref]

774. Cornelia Ferner, Werner Pomwenger, Stefan Wegenkittl, Martin Schnöll, Veronika Haaf, Arnold Keller. Information Extraction Engine for Sentiment-Topic Matching in Product Intelligence Applications 53-57. [Crossref]

775. Julie Dequaire, Peter Ondrúška, Dushyant Rao, Dominic Wang, Ingmar Posner. 2017. Deep tracking in the wild: End-to-end tracking using recurrent neural networks. *The International Journal of Robotics Research* 027836491771054. [Crossref]

776. Nathaniel Lageman, Eric D. Kilmer, Robert J. Walls, Patrick D. McDaniel. BinDNN: Resilient Function Matching Using Deep Learning 517-537. [Crossref]

777. Natthawut Kertkeidkachorn, Ryutaro Ichise. Estimating Distributed Representations of Compound Words Using Recurrent Neural Networks 235-246. [Crossref]

778. Ivan Kholod, Maria Efimova, Andrey Rukavitsyn, Shorov Andrey. Time Series Distributed Analysis in IoT with ETL and Data Mining Technologies 97-108. [Crossref]

779. K.G. Srinivasa, Sriram Anupindi, R. Sharath, S. Krishna Chaitanya. Analysis of Facial Expressiveness Captured in Reaction to Videos 664-670. [Crossref]

780. Marc Bolaños, Álvaro Peris, Francisco Casacuberta, Petia Radeva. VIBIKNet: Visual Bidirectional Kernelized Network for Visual Question Answering 372-380. [Crossref]

781. Emanuele Cipolla, Ignazio Infantino, Umberto Maniscalco, Giovanni Pilato, Filippo Vella. Indoor Actions Classification Through Long Short Term Memory Neural Networks 435-444. [Crossref]

782. Mohsen Fayyaz, Mohammad Hajizadeh Saffar, Mohammad Sabokrou, Mahmood Fathy, Fay Huang, Reinhard Klette. STFCN: Spatio-Temporal Fully Convolutional Neural Network for Semantic Segmentation of Street Scenes 493-509. [Crossref]

783. Fady Medhat, David Chesmore, John Robinson. Environmental Sound Recognition Using Masked Conditional Neural Networks 373-385. [Crossref]

784. Matthew Riemer. Distributed Computing in Social Media Analytics 121-135. [Crossref]

785. Yurun Shen, Hongtao Lu, Jie Jia. Classification of Motor Imagery EEG Signals with Deep Learning Models 181-190. [Crossref]

786. Steven Farrell, Dustin Anderson, Paolo Calafiura, Giuseppe Cerati, Lindsey Gray, Jim Kowalkowski, Mayur Mudigonda, Prabhat, Panagiotis Spentzouris, Maria Spiropoulou, Aristeidis Tsaris, Jean-Roch Vlimant, Stephan Zheng. 2017. The HEP.TrkX Project: deep neural networks for HL-LHC online and offline tracking. *EPJ Web of Conferences* **150**, 00003. [Crossref]

787. Daniele Ravi, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo, Guang-Zhong Yang. 2017. Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics* **21**:1, 4-21. [Crossref]

788. Rijun Liao, Chunshui Cao, Edel B. Garcia, Shiqi Yu, Yongzhen Huang. Pose-Based Temporal-Spatial Network (PTSN) for Gait Recognition with Carrying and Clothing Variations 474-483. [Crossref]

789. Lei Sun, Jun Du, Li-Rong Dai, Chin-Hui Lee. Multiple-target deep learning for LSTM-RNN based speech enhancement 136-140. [Crossref]

790. Alexandre Alahi, Vignesh Ramanathan, Kratarth Goel, Alexandre Robicquet, Amir A. Sadeghian, Li Fei-Fei, Silvio Savarese. Learning to Predict Human Behavior in Crowded Scenes 183-207. [Crossref]

791. Tianwen Jiang, Ming Liu, Bing Qin, Ting Liu. Constructing Semantic Hierarchies via Fusion Learning Architecture 136-148. [Crossref]

792. Stefan Ultes, Alexander Schmitt, Wolfgang Minker. Analysis of Temporal Features for Interaction Quality Estimation 367-379. [Crossref]

793. Yao Cheng, Chang Xu, Daisuke Mashima, Vrizlynn L. L. Thing, Yongdong Wu. PowerLSTM: Power Demand Forecasting Using Long Short-Term Memory Neural Network 727-740. [Crossref]

794. Daehyun Lee, Kyungsik Myung. Read my lips, login to the virtual world 434-435. [Crossref]

795. Hongli Deng, Lei Zhang, Lituan Wang. 2017. Global context-dependent recurrent neural network language model with sparse feature learning. *Neural Computing and Applications* . [Crossref]

796. Wei Qiao, Ying Li, Zhong-Hai Wu. 2017. DLTAP: A Network-efficient Scheduling Method for Distributed Deep Learning Workload in Containerized Cluster Environment. *ITM Web of Conferences* **12**, 03030. [Crossref]

797. Joerg Evermann, Jana-Rebecca Rehse, Peter Fettke. A Deep Learning Approach for Predicting Process Behaviour at Runtime 327-338. [Crossref]

798. Pedro A. B. Gomes, Eduardo C. de Carvalho, Helder M. Arruda, Paulo de Souza, Gustavo Pessin. Exploiting Recurrent Neural Networks in the Forecasting of Bees' Level of Activity 254-261. [Crossref]

799. Tao Xu, Yun Zhou. Fall Detection Based on Skeleton Data 475-483. [Crossref]

800. Yoshimitsu Aoki. 2017. Behavior Monitoring System by using Vision Sensors. *IEICE Communications Society Magazine* **11**:1, 30-38. [Crossref]

801. Dimos Makris, Maximos Kaliakatsos-Papakostas, Ioannis Karydis, Katia Lida Kermanidis. Combining LSTM and Feed Forward Neural Networks for Conditional Rhythm Composition 570-582. [Crossref]

802. Hamza Labbaci, Brahim Medjahed, Youcef Aklouf. A Deep Learning Approach for Long Term QoS-Compliant Service Composition 287-294. [Crossref]

803. Cedric De Boom, Rohan Agrawal, Samantha Hansen, Esh Kumar, Romain Yon, Ching-Wei Chen, Thomas Demeester, Bart Dhoedt. 2017. Large-scale user modeling with recurrent neural networks for music discovery on multiple time scales. *Multimedia Tools and Applications* . [Crossref]

804. Amin Khatami, Yonghang Tai, Abbas Khosravi, Lei Wei, Mohsen Moradi Dalvand, Min Zou, Saeid Nahavandi. A Deep Learning-Based Model for Tactile Understanding on Haptic Data Percutaneous Needle Treatment 317-325. [Crossref]

805. Abhinav Thanda, Shankar M. Venkatesan. Audio Visual Speech Recognition Using Deep Recurrent Neural Networks 98-109. [Crossref]

806. Hua Ma, Pierre Ambrosini, Theo van Walsum. Fast Prospective Detection of Contrast Inflow in X-ray Angiograms with Convolutional Neural Network and Recurrent Neural Network 453-461. [Crossref]

807. Harish Yenala, Manoj Chinnakotla, Jay Goyal. Convolutional Bi-directional LSTM for Detecting Inappropriate Query Suggestions in Web Search 3-16. [Crossref]

808. Can Luo, Di Yang, Jin Huang, Yang-Dong Deng. 2017. LSTM-Based Temperature Prediction for Hot-Axles of Locomotives. *ITM Web of Conferences* **12**, 01013. [Crossref]

809. Ningning Ma, Hai-Tao Zheng, Xi Xiao. An Ontology-Based Latent Semantic Indexing Approach Using Long Short-Term Memory Networks 185-199. [Crossref]

810. Guangxiao Song, Zhijie Wang, Fang Han, Shenyi Ding. Transfer Learning for Music Genre Classification 183-190. [Crossref]

811. Erick López, Carlos Valle, Héctor Allende, Esteban Gil. Efficient Training Over Long Short-Term Memory Networks for Wind Speed Forecasting 409-416. [Crossref]

812. Niek Tax, Ilya Verenich, Marcello La Rosa, Marlon Dumas. Predictive Business Process Monitoring with LSTM Neural Networks 477-492. [Crossref]

813. Chang'e Jia, Chengjie Sun, Bingquan Liu, Lei Lin. Combine Non-text Features with Deep Learning Structures Based on Attention-LSTM for Answer Selection 261-271. [Crossref]

814. Yijin Guan, Zhihang Yuan, Guangyu Sun, Jason Cong. FPGA-based accelerator for long short-term memory recurrent neural networks 629-634. [Crossref]

815. Pavel Golik, Zoltán Tüske, Kazuki Irie, Eugen Beck, Ralf Schlüter, Hermann Ney. The 2016 RWTH Keyword Search System for Low-Resource Languages 719-730. [Crossref]

816. Le-kui Zhou, Si-liang Tang, Jun Xiao, Fei Wu, Yue-ting Zhuang. 2017. Disambiguating named entities with deep supervised learning via crowd labels. *Frontiers of Information Technology & Electronic Engineering* **18**:1, 97-106. [Crossref]

817. Huiwei Zhou, Yunlong Yang, Zhuang Liu, Yingyu Lin, Pengfei Zhu, Degen Huang. Jointly Learning Bilingual Sentiment and Semantic Representations for Cross-Language Sentiment Classification 149-160. [Crossref]

818. Adebayo Kolawole John, Luigi Di Caro, Livio Robaldo, Guido Boella. Textual Inference with Tree-Structured LSTM 17-31. [Crossref]

819. Chengjie Sun, Yang Liu, Chang'e Jia, Bingquan Liu, Lei Lin. Recognizing Text Entailment via Bidirectional LSTM Model with Inner-Attention 448-457. [Crossref]

820. Aryo Pradipta Gema, Suhendro Winton, Theodorus David, Derwin Suhartono, Muhsin Shodiq, Wikaria Gazali. 2017. It Takes Two To Tango: Modification of Siamese Long Short Term Memory Network with Attention Mechanism in Recognizing Argumentative Relations in Persuasive Essay. *Procedia Computer Science* **116**, 449-459. [Crossref]

821. Ignacio Arnaldo, Alfredo Cuesta-Infante, Ankit Arun, Mei Lam, Costas Bassias, Kalyan Veeramachaneni. Learning Representations for Log Data in Cybersecurity 250-268. [Crossref]

822. Qing Li, Yun Ma, Zhenguo Yang. Event Cube – A Conceptual Framework for Event Modeling and Analysis 499-515. [Crossref]

823. Xing Wu, Zhikang Du, Mingyu Zhong, Shuji Dai, Yazhou Liu. Chinese Lyrics Generation Using Long Short-Term Memory Neural Network 419-427. [Crossref]

824. Nicha C. Dvornek, Pamela Ventola, Kevin A. Pelphrey, James S. Duncan. Identifying Autism from Resting-State fMRI Using Long Short-Term Memory Networks 362-370. [Crossref]

825. Najmeh Sadoughi, Carlos Busso. **10498**, 389. [Crossref]

826. Phyo P. San, Pravin Kakar, Xiao-Li Li, Shonali Krishnaswamy, Jian-Bo Yang, Minh N. Nguyen. Deep Learning for Human Activity Recognition 186-204. [Crossref]

827. Yuntian Feng, Hongjun Zhang, Wenning Hao, Gang Chen. 2017. Joint Extraction of Entities and Relations Using Reinforcement Learning and Deep Learning. *Computational Intelligence and Neuroscience* **2017**, 1-11. [Crossref]

828. Yuki Endo, Kyosuke Nishida, Hiroyuki Toda, Hiroshi Sawada. Predicting Destinations from Partial Trajectories Using Recurrent Neural Network 160-172. [Crossref]

829. Wei Wang, Sergey Tulyakov, Nicu Sebe. Recurrent Convolutional Face Alignment 104-120. [Crossref]

830. Chonho Lee, Zhaojing Luo, Kee Yuan Ngiam, Meihui Zhang, Kaiping Zheng, Gang Chen, Beng Chin Ooi, Wei Luen James Yip. Big Healthcare Data Analytics: Challenges and Applications 11-41. [Crossref]

831. Xiaoxu Li, Chensi Mao, Shiliang Huang, Zhongfu Ye. Chinese Sign Language Recognition Based on SHS Descriptor and Encoder-Decoder LSTM Model 719-728. [Crossref]

832. Zhuang Liu, Degen Huang, Yunxia Zhang, Chuan Zhang. A Hierarchical Iterative Attention Model for Machine Comprehension 341-352. [Crossref]

833. Markus Müller, Sebastian Stüker, Alex Waibel. Language Adaptive Multilingual CTC Speech Recognition 473-482. [Crossref]

834. Michal Sofka, Fausto Milletari, Jimmy Jia, Alex Rothberg. Fully Convolutional Regression Network for Accurate Detection of Measurement Points 258-266. [Crossref]

835. Yong-hong Tian, Xi-lin Chen, Hong-kai Xiong, Hong-liang Li, Li-rong Dai, Jing Chen, Jun-liang Xing, Jing Chen, Xi-hong Wu, Wei-min Hu, Yu Hu, Tie-jun Huang, Wen Gao. 2017. Towards human-like and transhuman perception in AI 2.0: a review. *Frontiers of Information Technology & Electronic Engineering* **18**:1, 58-67. [Crossref]

836. Tao Zhang, Qing Li, Chang-shui Zhang, Hua-wei Liang, Ping Li, Tian-miao Wang, Shuo Li, Yun-long Zhu, Cheng Wu. 2017. Current trends in the development of intelligent unmanned autonomous systems. *Frontiers of Information Technology & Electronic Engineering* **18**:1, 68-85. [Crossref]

837. Victoria Ruiz, Maria T. Gonzalez de Lena, Jorge Sueiras, Angel Sanchez, Jose F. Velez. A Deep Learning Approach to Handwritten Number Recognition 193-202. [Crossref]

838. Zizhao Zhang, Pingjun Chen, Manish Sapkota, Lin Yang. TandemNet: Distilling Knowledge from Medical Images Using Diagnostic Reports as Optional Semantic References 320-328. [Crossref]

839. Lin Wu, Michele Haynes, Andrew Smith, Tong Chen, Xue Li. Generating Life Course Trajectory Sequences with Recurrent Neural Networks and Application to Early Detection of Social Disadvantage 225-242. [Crossref]

840. David Greenwood, Stephen Laycock, Iain Matthews. **10498**, 160. [Crossref]

841. Chi Thang Duong, Quoc Viet Hung Nguyen, Sen Wang, Bela Stantic. Provenance-Based Rumor Detection 125-137. [Crossref]

842. Xiang Li, Rui Yan, Ming Zhang. Joint Emoji Classification and Embedding Learning 48-63. [Crossref]

843. Ruochun Jin, Jingfei Jiang, Yong Dou. Accuracy Evaluation of Long Short Term Memory Network Based Language Model with Fixed-Point Arithmetic 281-288. [Crossref]

844. Jinkyu Lee, Keulbit Kim, Turaj Shabestary, Hong-Goo Kang. Deep bi-directional long short-term memory based speech enhancement for wind noise reduction 41-45. [Crossref]

845. Xilian Li, Wei Chen, Tengjiao Wang, Weijing Huang. Target-Specific Convolutional Bi-directional LSTM Neural Network for Political Ideology Analysis 64-72. [Crossref]

846. Rudra P. K. Poudel, Pablo Lamata, Giovanni Montana. Recurrent Fully Convolutional Neural Networks for Multi-slice MRI Cardiac Segmentation 83-94. [Crossref]

847. Tseng-Hung Chen, Kuo-Hao Zeng, Wan-Ting Hsu, Min Sun. Video Captioning via Sentence Augmentation and Spatio-Temporal Attention 269-286. [Crossref]

848. Deepika Singh, Erinc Merdivan, Ismini Psychoula, Johannes Kropf, Sten Hanke, Matthieu Geist, Andreas Holzinger. Human Activity Recognition Using Recurrent Neural Networks 267-274. [Crossref]

849. Yu-Hsiang Huang, Hen-Hsen Huang, Hsin-Hsi Chen. Irony Detection with Attentive Recurrent Neural Networks 534-540. [Crossref]

850. Chi Lu, Heyan Huang, Ping Jian, Dan Wang, Yi-Di Guo. A P-LSTM Neural Network for Sentiment Classification 524-533. [Crossref]

851. Yuhan Jia, Jianping Wu, Ming Xu. 2017. Traffic Flow Prediction with Rainfall Impact Using a Deep Learning Method. *Journal of Advanced Transportation* **2017**, 1-10. [Crossref]

852. Shrooq Alsenan, Mourad Ykhlef. Statistical Machine Translation Context Modelling with Recurrent Neural Network and LDA 75-84. [Crossref]

853. Sonam Singh, Sudeshna Sarkar, Pabitra Mitra. Leveraging Convolutions in Recurrent Neural Networks for Doppler Weather Radar Echo Prediction 310-317. [Crossref]

854. Suraj Srinivas, Ravi K. Sarvadevabhatla, Konda R. Mopuri, Nikita Prabhu, Srinivas S.S. Kruthiventi, R. Venkatesh Babu. An Introduction to Deep Convolutional Neural Nets for Computer Vision 25-52. [Crossref]

855. Shruthi Gopalswamy, Patrick J. Tighe, Parisa Rashidi. Deep recurrent neural networks for predicting intraoperative and postoperative outcomes and trends 361-364. [Crossref]

856. Aitor Almeida, Gorka Azkune. Inter-activity Behaviour Modelling Using Long Short-Term Memory Networks 394-399. [Crossref]

857. Di Cao, Kai Yu. Deep Attentive Structured Language Model Based on LSTM 169-180. [Crossref]

858. Amir Rosenfeld, Shimon Ullman. Visual Concept Recognition and Localization via Iterative Introspection 264-279. [Crossref]

859. Liangfu Cao, Lianli Gao, Jingkuan Song, Xing Xu, Heng Tao Shen. Jointly Learning Attentions with Semantic Cross-Modal Correlation for Visual Question Answering 248-260. [Crossref]

860. Tian Tian, WeiRan Xu. Chinese Lexical Normalization Based on Information Extraction: An Experimental Study 216-223. [Crossref]

861. Yao Lu, Chunyun Zhang, Weiran Xu. Instance-Adaptive Attention Mechanism for Relation Classification 322-330. [Crossref]

862. Fady Medhat, David Chesmore, John Robinson. Masked Conditional Neural Networks for Audio Classification 349-358. [Crossref]

863. Chandrakant Bothe, Sven Magg, Cornelius Weber, Stefan Wermter. Dialogue-Based Neural Learning to Estimate the Sentiment of a Next Upcoming Utterance 477-485. [Crossref]

864. Andrew D. Mann, Denise Gorse. A New Methodology to Exploit Predictive Power in (Open, High, Low, Close) Data 495-502. [Crossref]

865. Stefan Reitmann, Karl Nachtigall. Applying Bidirectional Long Short-Term Memories (BLSTM) to Performance Data in Air Traffic Management for System Identification 528-536. [Crossref]

866. Kebin Peng, Dianfu Ma. Tree-Structure CNN for Automated Theorem Proving 3-12. [Crossref]

867. Weijie An, Qin Chen, Yan Yang, Liang He. Knowledge Memory Based LSTM Model for Answer Selection 34-42. [Crossref]

868. Qianli Ma, Lifeng Shen, Ruishi Su, Jieyu Chen. Two-Stage Temporal Multimodal Learning for Speaker and Speech Recognition 64-72. [Crossref]

869. Jizhao Zhu, Jianzhong Qiao, Xinxiao Dai, Xueqi Cheng. Relation Classification via Target-Concentrated Attention CNNs 137-146. [Crossref]

870. Nicolas Granger, Mounîm A. el Yacoubi. Comparing Hybrid NN-HMM and RNN for Temporal Modeling in Gesture Recognition 147-156. [Crossref]

871. Rustem Takhanov, Zhenisbek Assylbekov. Patterns Versus Characters in Subword-Aware Neural Language Modeling 157-166. [Crossref]

872. Xiaolei Niu, Yuexian Hou. Hierarchical Attention BLSTM for Modeling Sentences and Documents 167-177. [Crossref]

873. Xiaolei Niu, Yuexian Hou, Panpan Wang. Bi-Directional LSTM with Quantum Attention Mechanism for Sentence Modeling 178-188. [Crossref]

874. Shuanglong Liu, Chao Zhang, Jinwen Ma. CNN-LSTM Neural Network Model for Quantitative Strategy Analysis in Stock Markets 198-206. [Crossref]

875. Jegyung Son, Gil-Jin Jang, Minho Lee. Temporal Attention Neural Network for Video Understanding 422-430. [Crossref]

876. Anupiya Nugaliyadde, Kok Wai Wong, Ferdous Sohel, Hong Xie. Reinforced Memory Network for Question Answering 482-490. [Crossref]

877. Sadam Al-Azani, El-Sayed M. El-Alfy. Hybrid Deep Learning for Sentiment Polarity Determination of Arabic Microblogs 491-500. [Crossref]

878. Fengyu Li, Mingmin Chi, Dong Wu, Junyu Niu. Hierarchical Parameter Sharing in Recursive Neural Networks with Long Short-Term Memory 582-592. [Crossref]

879. Lap Pong Cheung, Haiqin Yang. Heterogeneous Features Integration in Deep Knowledge Tracing 653-662. [Crossref]

880. Xiaokang Tu, Yuchun Fang. Ultra-deep Neural Network for Face Anti-spoofing 686-695. [Crossref]

881. Jiayi Zhang, Wenge Rong, Qiubin Liang, Haonan Sun, Zhang Xiong. Data Augmentation Based Stock Trend Prediction Using Self-organising Map 903-912. [Crossref]

882. Andrew D. Mann, Denise Gorse. Deep Candlestick Mining 913-921. [Crossref]

883. Zhonglei Lu, Wenfen Liu, Yanfang Zhou, Xuexian Hu, Binyu Wang. An Effective Approach of Sentence Compression Based on "Re-read" Mechanism and Bayesian Combination Model 129-140. [Crossref]

884. Suyang Zhu, Shoushan Li, Guodong Zhou. Opinion Target Understanding in Event-Level Sentiment Analysis 153-165. [Crossref]

885. Jian Xu, Hao Yin, Shoushan Li, Guodong Zhou. Supervised Domain Adaptation for Sentiment Regression 193-205. [Crossref]

886. Hongjian Zhan, Qingqing Wang, Yue Lu. Handwritten Digit String Recognition by Combination of Residual Network and RNN-CTC 583-591. [Crossref]

887. Yuanyuan Zhao, Shiyu Zhou, Shuang Xu, Bo Xu. Word-Level Permutation and Improved Lower Frame Rate for RNN-Based Acoustic Modeling 859-869. [Crossref]

888. Yujie Wang, Biao Leng, Guanglu Song. Spatial Quality Aware Network for Video-Based Person Re-identification 34-43. [Crossref]

889. Bo Huang, Hualong Huang, Hongtao Lu. Convolutional Gated Recurrent Units Fusion for Video Action Recognition 114-123. [Crossref]

890. Anthony F. Botelho, Ryan S. Baker, Neil T. Heffernan. Improving Sensor-Free Affect Detection Using Deep Learning 40-51. [Crossref]

891. Jingyuan Wang, Fei Hu, Li Li. Deep Bi-directional Long Short-Term Memory Model for Short-Term Traffic Flow Prediction 306-316. [Crossref]

892. Ke Wang, Peidong Zhu, Haoyang Zhu, Pengshuai Cui, Zhenyu Zhang. An Interweaved Time Series Locally Connected Recurrent Neural Network Model on Crime Forecasting 466-474. [Crossref]

893. Yagmur Gizem Cinar, Hamid Mirisaee, Parantapa Goswami, Eric Gaussier, Ali Aït-Bachir, Vadim Strijov. Position-Based Content Attention for Time Series Forecasting with Sequence-to-Sequence RNNs 533-544. [Crossref]

894. Noëlie Cherrier, Thibaut Castaings, Alexandre Boulch. Deep Sequence-to-Sequence Neural Networks for Ionospheric Activity Map Prediction 545-555. [Crossref]

895. Wei Lee Woon, Stefan Oehmcke, Oliver Kramer. Spatio-Temporal Wind Power Prediction Using Recurrent Neural Networks 556-563. [Crossref]

896. Hui Song, A. K. Qin, Flora D. Salim. Multi-resolution Selective Ensemble Extreme Learning Machine for Electricity Consumption Prediction 600-609. [Crossref]

897. Dehua Chen, Cheng Shi, Mei Wang, Qiao Pan. Thyroid Nodule Classification Using Hierarchical Recurrent Neural Network with Multiple Ultrasound Reports 765-773. [Crossref]

898. Amin Khatami, Yonghang Tai, Abbas Khosravi, Lei Wei, Mohsen Moradi Dalvand, Jun Peng, Saeid Nahavandi. A Haptics Feedback Based-LSTM Predictive Model for Pericardiocentesis Therapy Using Public Intreoperative Data 810-818. [Crossref]

899. Aniruddha Parvat, Jai Chavan, Siddhesh Kadam, Souradeep Dev, Vidhi Pathak. A survey of deep-learning frameworks 1-7. [Crossref]

900. Deepika Singh, Erinc Merdivan, Sten Hanke, Johannes Kropf, Matthieu Geist, Andreas Holzinger. Convolutional and Recurrent Neural Networks for Activity Recognition in Smart Environment 194-205. [Crossref]

901. Artem M. Grachev, Dmitry I. Ignatov, Andrey V. Savchenko. Neural Networks Compression for Language Modeling 351-357. [Crossref]

902. Prateep Bhattacharjee, Sukhendu Das. Two-Stream Convolutional Network with Multi-level Feature Fusion for Categorization of Human Action from Videos 549-556. [Crossref]

903. Hakan Erdogan, John R. Hershey, Shinji Watanabe, Jonathan Le Roux. Deep Recurrent Networks for Separation and Recognition of Single-Channel Speech in Nonstationary Background Audio 165-186. [Crossref]

904. Yu Zhang, Dong Yu, Guoguo Chen. Advanced Recurrent Neural Networks for Automatic Speech Recognition 261-279. [Crossref]

905. Yajie Miao, Florian Metze. End-to-End Architectures for Speech Recognition 299-323. [Crossref]

906. Yifan Gong, Yan Huang, Kshitiz Kumar, Jinyu Li, Chaojun Liu, Guoli Ye, Shixiong Zhang, Yong Zhao, Rui Zhao. Challenges in and Solutions to Deep Learning Network Acoustic Modeling in Speech Recognition Products at Microsoft 401-417. [Crossref]

907. Danilo Croce, Simone Filice, Roberto Basili. On the Impact of Linguistic Information in Kernel-Based Deep Architectures 359-371. [Crossref]

908. Wenwen Ye, Yongfeng Zhang, Wayne Xin Zhao, Xu Chen, Zheng Qin. A Collaborative Neural Model for Rating Prediction by Leveraging User Reviews and Product Images 99-111. [Crossref]

909. Filippo Maria Bianchi, Enrico Maiorino, Michael C. Kampffmeyer, Antonello Rizzi, Robert Jenssen. Introduction 1-7. [Crossref]

910. Filippo Maria Bianchi, Enrico Maiorino, Michael C. Kampffmeyer, Antonello Rizzi, Robert Jenssen. Recurrent Neural Network Architectures 23-29. [Crossref]

911. Rafael T. Gonzalez, Jaime A. Riascos, Dante A. C. Barone. How Artificial Intelligence is Supporting Neuroscience Research: A Discussion About Foundations, Methods and Applications 63-77. [Crossref]

912. Mohamed Seghir Hadj Ameur, Farid Meziane, Ahmed Guessoum. 2017. Arabic Machine Transliteration using an Attention-based Encoder-decoder Model. *Procedia Computer Science* 117, 287-297. [Crossref]

913. Mattia Antonino Di Gangi, Salvatore Gaglio, Claudio La Bua, Giosué Lo Bosco, Riccardo Rizzo. A Deep Learning Network for Exploiting Positional Information in Nucleosome Related Sequences 524-533. [Crossref]

914. Fuxue Li, Du Quan, Wang Qiang, Xiao Tong, Jingbo Zhu. Handling Many-To-One UNK Translation for Neural Machine Translation 102-111. [Crossref]

915. Yunxiang Mao, Zhaozheng Yin. Two-Stream Bidirectional Long Short-Term Memory for Mitosis Event Detection and Stage Localization in Phase-Contrast Microscopy Images 56-64. [Crossref]

916. Bo Zheng, Wanxiang Che, Jiang Guo, Ting Liu. Enhancing LSTM-based Word Segmentation Using Unlabeled Data 60-70. [Crossref]

917. Yang Zhao, Yining Wang, Jiajun Zhang, Chengqing Zong. Cost-Aware Learning Rate for Neural Machine Translation 85-93. [Crossref]

918. Chunqi Wang, Wei Chen, Bo Xu. Named Entity Recognition with Gated Convolutional Neural Networks 110-121. [Crossref]

919. Peng Zhou, Suncong Zheng, Jiaming Xu, Zhenyu Qi, Hongyun Bao, Bo Xu. Joint Extraction of Multiple Relations and Entities by Using a Hybrid Neural Network 135-146. [Crossref]

920. Wei Wu, Houfeng Wang, Sujian Li. Bi-directional Gated Memory Networks for Answer Selection 251-262. [Crossref]

921. Run-Ze Wang, Chen-Di Zhan, Zhen-Hua Ling. Question Answering with Character-Level LSTM Encoders and Model-Based Data Augmentation 295-305. [Crossref]

922. Xinqi Bao, Yunfang Wu. Exploiting Explicit Matching Knowledge with Long Short-Term Memory 306-317. [Crossref]

923. Yidi Guo, Heyan Huang, Yang Gao, Chi Lu. Conceptual Multi-layer Neural Network Model for Headline Generation 355-367. [Crossref]

924. Zhuang Liu, Degen Huang, Kaiyu Huang, Jing Zhang. DIM Reader: Dual Interaction Model for Machine Comprehension 387-397. [Crossref]

925. Yue Wu, Tianxing He, Zhehuai Chen, Yanmin Qian, Kai Yu. Multi-view LSTM Language Model with Word-Synchronized Auxiliary Feature for LVCSR 398-410. [Crossref]

926. Yang Liu, Jiajun Zhang, Chengqing Zong. Memory Augmented Attention Model for Chinese Implicit Discourse Relation Recognition 411-423. [Crossref]

927. Nuo Qun, Xing Li, Xipeng Qiu, Xuanjing Huang. End-to-End Neural Text Classification for Tibetan 472-480. [Crossref]

928. Sumit Gulwani, Prateek Jain. Programming by Examples: PL Meets ML 3-20. [Crossref]

929. Pak-Kan Wong, Man-Leung Wong, Kwong-Sak Leung. Long-Short Term Memory Network for RNA Structure Profiling Super-Resolution 255-266. [Crossref]

930. Fady Medhat, David Chesmore, John Robinson. Masked Conditional Neural Networks for Environmental Sound Classification 21-33. [Crossref]

931. Sascha Krstanovic, Heiko Paulheim. Ensembles of Recurrent Neural Networks for Robust Time Series Forecasting 34-46. [Crossref]

932. Yu Wang, Yang Feng, Zhe Hong, Ryan Berger, Jiebo Luo. How Polarized Have We Become? A Multimodal Classification of Trump Followers and Clinton Followers 440-456. [Crossref]

933. Zheng Huang, Jiajun Peng, Huijuan Lian, Jie Guo, Weidong Qiu. 2017. Deep Recurrent Model for Server Load and Performance Prediction in Data Center. *Complexity* **2017**, 1-10. [Crossref]

934. Harika Abburi, Rajendra Prasath, Manish Shrivastava, Suryakanth V. Gangashetty. Significance of DNN-AM for Multimodal Sentiment Analysis 231-239. [Crossref]

935. S. Sachin Kumar, M. Anand Kumar, K. P. Soman. Sentiment Analysis of Tweets in Malayalam Using Long Short-Term Memory Units and Convolutional Neural Nets 320-334. [Crossref]

936. Nga Nguyen Thi, Van Loi Cao, Nhien-An Le-Khac. One-Class Collective Anomaly Detection Based on LSTM-RNNs 73-85. [Crossref]

937. Chensi Mao, Shiliang Huang, Xiaoxu Li, Zhongfu Ye. Chinese Sign Language Recognition with Sequence to Sequence Learning 180-191. [Crossref]

938. Chunlin Tian, Yuan Yuan, Xiaoqiang Lu. Deep Temporal Architecture for Audiovisual Speech Recognition 650-661. [Crossref]

939. Yan Sun, Bo Ren. Automatic Image Description Generation with Emotional Classifiers 748-763. [Crossref]

940. Yonghao Xu, Bo Du, Liangpei Zhang, Fan Zhang. A Band Grouping Based LSTM Algorithm for Hyperspectral Image Classification 421-432. [Crossref]

941. Yiran Chen, Hai (Helen) Li, Chunpeng Wu, Chang Song, Sicheng Li, Chuhan Min, Hsin-Pai Cheng, Wei Wen, Xiaoxiao Liu. 2017. Neuromorphic computing's yesterday, today, and tomorrow – an evolutional view. *Integration, the VLSI Journal* . [Crossref]

942. Ayanava Sarkar, Alexander Gepperth, Uwe Handmann, Thomas Kopinski. Dynamic Hand Gesture Recognition for Mobile Systems Using Deep LSTM 19-31. [Crossref]

943. Ze Hu, Zhan Zhang, Haiqin Yang, Qing Chen, Rong Zhu, Decheng Zuo. 2017. Predicting the quality of online health expert question-answering services with temporal features in a deep learning framework. *Neurocomputing* . [Crossref]

944. Yuedong Yang, Jianzhao Gao, Jihua Wang, Rhys Heffernan, Jack Hanson, Kuldip Paliwal, Yaoqi Zhou. 2016. Sixty-five years of the long march in protein secondary structure prediction: the final stretch?. *Briefings in Bioinformatics* bbw129. [Crossref]

945. Franck Dernoncourt, Ji Young Lee, Ozlem Uzuner, Peter Szolovits. 2016. De-identification of patient notes with recurrent neural networks. *Journal of the American Medical Informatics Association* ocw156. [Crossref]

946. Jack Hanson, Yuedong Yang, Kuldip Paliwal, Yaoqi Zhou. 2016. Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks. *Bioinformatics* btw678. [Crossref]

947. Peratham Wiriyathammabhum, Douglas Summers-Stay, Cornelia Fermüller, Yiannis Aloimonos. 2016. Computer Vision and Natural Language Processing. *ACM Computing Surveys* **49**:4, 1-44. [Crossref]

948. Gregory B Cogan, Asha Iyer, Lucia Melloni, Thomas Thesen, Daniel Friedman, Werner Doyle, Orrin Devinsky, Bijan Pesaran. 2016. Manipulating stored phonological input during verbal working memory. *Nature Neuroscience* **20**:2, 279-286. [Crossref]

949. Youngsoo Jang, Jiyeon Ham, Byung-Jun Lee, Youngjae Chang, Kee-Eung Kim. Neural dialog state tracker for large ontologies by attention mechanism 531-537. [Crossref]

950. Hamid Reza Hassanzadeh, May D. Wang. DeeperBind: Enhancing prediction of sequence specificities of DNA binding proteins 178-183. [Crossref]

951. Meng Joo Er, Yong Zhang, Ning Wang, Mahardhika Pratama. 2016. Attention pooling-based convolutional neural network for sentence modelling. *Information Sciences* **373**, 388-403. [Crossref]

952. Takaaki Hori, Hai Wang, Chiori Hori, Shinji Watanabe, Bret Harsham, Jonathan Le Roux, John R. Hershey, Yusuke Koji, Yi Jing, Zhaocheng Zhu, Takeyuki Aikawa. Dialog state tracking with attention-based sequence-to-sequence learning 552-558. [Crossref]

953. Lyndon White, Roberto Togneri, Wei Liu, Mohammed Bennamoun. Modelling Sentence Generation from Sum of Word Embedding Vectors as a Mixed Integer Programming Problem 770-777. [Crossref]

954. Tu-Hoa Pham, Adrien Bufort, Stephane Caron, Abderrahmane Kheddar. Whole-body contact force sensing from motion capture 58-63. [Crossref]

955. Siyu Huang, Xi Li, Zhongfei Zhang, Zhouzhou He, Fei Wu, Wei Liu, Jinhui Tang, Yueting Zhuang. 2016. Deep Learning Driven Visual Path Prediction From a Single Image. *IEEE Transactions on Image Processing* **25**:12, 5892-5904. [Crossref]

956. Firoj Alam, Shammur Absar Chowdhury, Sheak Rashed Haider Noori. Bidirectional LSTMs — CRFs networks for bangla POS tagging 377-382. [Crossref]

957. Yanmin Qian, Nanxin Chen, Kai Yu. 2016. Deep features for automatic spoofing detection. *Speech Communication* **85**, 43-52. [Crossref]

958. Yun-Nung Chen, Dilek Hakanni-Tur, Gokhan Tur, Asli Celikyilmaz, Jianfeng Guo, Li Deng. Syntax or semantics? knowledge-guided joint semantic frame parsing 348-355. [Crossref]

959. Hamid Palangi, Rabab Ward, Li Deng. Reconstruction of sparse vectors in compressive sensing with multiple measurement vectors using bidirectional long short-term memory 192-196. [Crossref]

960. Ke Zhang, Jianwu Xu, Martin Renqiang Min, Guofei Jiang, Konstantinos Pelechrinis, Hui Zhang. Automated IT system failure prediction: A deep learning approach 1291-1300. [Crossref]

961. Anqi Li, Michael Lewis, Christian Lebiere, Katia Sycara, Shehzaman Salim Khatib, Yuqing Tang, Matthew Siedsma, Don Morrison. A computational model based on human performance for fluid management in critical care 1-6. [Crossref]

962. Junnan Li, Yongkang Wong, Mohan S. Kankanhalli. Multi-stream Deep Learning Framework for Automated Presentation Assessment 222-225. [Crossref]

963. Ondrej Klejch, Peter Bell, Steve Renals. Punctuated transcription of multi-genre broadcasts using acoustic and lexical approaches 433-440. [Crossref]

964. Wei Fang, Juei-Yang Hsu, Hung-yi Lee, Lin-Shan Lee. Hierarchical attention model for improved machine comprehension of spoken content 232-238. [Crossref]

965. Guoqiang Zhong, Li-Na Wang, Xiao Ling, Junyu Dong. 2016. An overview on data representation learning: From traditional feature learning to recent deep learning. *The Journal of Finance and Data Science* **2**:4, 265-278. [Crossref]

966. Mohamed Bouaziz, Mohamed Morchid, Richard Dufour, Georges Linares. Improving multi-stream classification by mapping sequence-embedding in a high dimensional space 224-231. [Crossref]

967. Nastaran Mohammadian Rad, Seyed Mostafa Kia, Calogero Zarbo, Giuseppe Jurman, Paola Venuti, Cesare Furlanello. Stereotypical Motor Movement Detection in Dynamic Feature Space 487-494. [Crossref]

968. Yujin Tang, Jianfeng Xu, Kazunori Matsumoto, Chihiro Ono. Sequence-to-Sequence Model with Attention for Time Series Classification 503-510. [Crossref]

969. Mohamed Elfeky, Meysam Bastani, Xavier Velez, Pedro Moreno, Austin Waters. Towards acoustic model unification across dialects 624-628. [Crossref]

970. Aleksandr Sboev, Tatiana Litvinova, Irina Voronina, Dmitry Gudovskikh, Roman Rybka. Deep Learning Network Models to Categorize Texts According to Author's Gender and to Identify Text Sentiment 1101-1106. [Crossref]

971. Shane Settle, Karen Livescu. Discriminative acoustic word embeddings: Tecurrent neural network-based approaches 503-510. [Crossref]

972. Asif Hassan, Mohammad Rashedul Amin, Abul Kalam Al Azad, Nabeel Mohammed. Sentiment analysis on bangla and romanized bangla text using deep recurrent models 51-56. [Crossref]

973. Geetika Garg, Chris Pollett. Neural network CAPTCHA crackers 853-861. [Crossref]

974. Mohamed Kamal Omar. A factor analysis model of sequences for language recognition 341-347. [Crossref]

975. Dario Bertero, Pascale Fung. Multimodal deep neural nets for detecting humor in TV sitcoms 383-390. [Crossref]

976. Dickson Neoh Tze How, Chu Kiong Loo, Khairul Salleh Mohamed Sahari. 2016. Behavior recognition for humanoid robots using long short-term memory. *International Journal of Advanced Robotic Systems* **13**:6, 172988141666336. [Crossref]

977. Takashi Ushio, Hongjie Shi, Mitsuru Endo, Katsuyoshi Yamagami, Noriaki Horii. Recurrent convolutional neural networks for structured speech act tagging 518-524. [Crossref]

978. Shubham Toshniwal, Karen Livescu. Jointly learning to align and convert graphemes to phonemes with neural attention models 76-82. [Crossref]

979. Yinglin Wang, Jianzhang Zhang. Experiment on automatic functional requirements analysis with the EFRF's semantic cases 636-642. [Crossref]

980. Niek Tax, Natalia Sidorova, Reinder Haakma, Wil M.P. van der Aalst. 2016. Mining local process models. *Journal of Innovation in Digital Ecosystems* **3**:2, 183-196. [Crossref]

981. Tian Gao, Jun Du, Yong Xu, Cong Liu, Li-Rong Dai, Chin-Hui Lee. 2016. Joint training of DNNs by incorporating an explicit dereverberation structure for distant speech recognition. *EURASIP Journal on Advances in Signal Processing* **2016**:1. . [Crossref]

982. Daniel L. Marino, Kasun Amarasinghe, Milos Manic. Simultaneous generation-classification using LSTM 1-8. [Crossref]

983. Rajendra Rana Bhat, Rodrigo Daniel Trevizan, Rahul Sengupta, Xiaolin Li, Arturo Bretas. Identifying Nontechnical Power Loss via Spatial and Temporal Deep Learning 272-279. [Crossref]

984. Shimul Hassan, Nabeel Mohammed, Sifat Momen. Learning from tagore 79-84. [Crossref]

985. Mohamed Bouaziz, Mohamed Morchid, Richard Dufour, Georges Linares, Renato De Mori. Parallel Long Short-Term Memory for multi-stream classification 218-223. [Crossref]

986. Joo-Kyung Kim, Gokhan Tur, Asli Celikyilmaz, Bin Cao, Ye-Yi Wang. Intent detection using semantically enriched word embeddings 414-419. [Crossref]

987. Chao Yuan, Amit Chakraborty. Deep Convolutional Factor Analyser for Multivariate Time Series Modeling 1323-1328. [Crossref]

988. Ming Sun, Anirudh Raju, George Tucker, Sankaran Panchapagesan, Gengshen Fu, Arindam Mandal, Spyros Matsoukas, Nikko Strom, Shiv Vitaladevuni. Max-pooling loss training of long short-term memory networks for small-footprint keyword spotting 474-480. [Crossref]

989. Kai Zhao, Sasu Tarkoma, Siyuan Liu, Huy Vo. Urban human mobility data mining: An overview 1911-1920. [Crossref]

990. Morten Kolboek, Zheng-Hua Tan, Jesper Jensen. Speech enhancement using Long Short-Term Memory based recurrent Neural Networks for noise robust Speaker Verification 305-311. [Crossref]

991. Hoang Minh Nguyen, Sungpil Woo, Janggwan Im, Taejoon Jun, Daeyoung Kim. A Workload Prediction Approach Using Models Stacking Based on Recurrent Neural Network and Autoencoder 929-936. [Crossref]

992. Chaochun Liu, Huan Sun, Nan Du, Shulong Tan, Hongliang Fei, Wei Fan, Tao Yang, Hao Wu, Yaliang Li, Chenwei Zhang. Augmented LSTM Framework to Construct Medical Self-Diagnosis Android 251-260. [Crossref]

993. Andrew S. Davis, Itamar Arel. Faster Gated Recurrent Units via Conditional Computation 920-924. [Crossref]

994. Shuhan Yuan, Xintao Wu, Yang Xiang. Incorporating Pre-Training in Long Short-Term Memory Networks for Tweets Classification 1329-1334. [Crossref]

995. Lizhe Liu, Weiting Chen, Guitao Cao. Prediction of neonatal amplitude-integrated EEG based on LSTM method 497-500. [Crossref]

996. Lang-Chi Yu, Hung-yi Lee, Lin-shan Lee. Abstractive headline generation for spoken content by attentive recurrent neural networks with ASR error modeling 151-157. [Crossref]

997. Lina Maria Rojas-Barahona. 2016. Deep learning for sentiment analysis. *Language and Linguistics Compass* **10**:12, 701-719. [Crossref]

998. Uwe Springmann. 2016. OCR für alte Drucke. *Informatik-Spektrum* **39**:6, 459-462. [Crossref]

999. Scott Novotney, Damianos Karakos, Jan Silovsky, Rich Schwartz. BBN technologies' OpenSAD system 8-12. [Crossref]

1000. Wei-Ning Hsu, Yu Zhang, James Glass. A prioritized grid long short-term memory RNN for speech recognition 467-473. [Crossref]

1001. Dapeng Tao, Yonggang Wen, Richang Hong. 2016. Multicolumn Bidirectional Long Short-Term Memory for Mobile Devices-Based Human Activity Recognition. *IEEE Internet of Things Journal* **3**:6, 1124-1134. [Crossref]

1002. Tomohiro Tanaka, Takafumi Moriya, Takahiro Shinozaki, Shinji Watanabe, Takaaki Hori, Kevin Duh. Automated structure discovery and parameter tuning of neural network language model based on evolution strategy 665-671. [Crossref]

1003. Tuka AlHanai, Wei-Ning Hsu, James Glass. Development of the MIT ASR system for the 2016 Arabic Multi-genre Broadcast Challenge 299-304. [Crossref]

1004. Du-Xin Liu, Wenbin Du, Xinyu Wu, Can Wang, Yu Qiao. Deep rehabilitation gait learning for modeling knee joints of lower-limb exoskeleton 1058-1063. [Crossref]

1005. Jia Wu, Weiru Zeng, Zhe Chen, Xue-Fei Tang. Hierarchical Temporal Memory Method for Time-Series-Based Anomaly Detection 1167-1172. [Crossref]

1006. Suncong Zheng, Jiaming Xu, Peng Zhou, Hongyun Bao, Zhenyu Qi, Bo Xu. 2016. A neural network framework for relation extraction: Learning entity semantic and relation pattern. *Knowledge-Based Systems* **114**, 12-23. [Crossref]

1007. Kai Xu, Yifei Shi, Lintao Zheng, Junyu Zhang, Min Liu, Hui Huang, Hao Su, Daniel Cohen-Or, Baoquan Chen. 2016. 3D attention-driven depth acquisition for object identification. *ACM Transactions on Graphics* **35**:6, 1-14. [Crossref]

1008. Yuwei Cui, Subutai Ahmad, Jeff Hawkins. 2016. Continuous Online Sequence Learning with an Unsupervised Neural Network Model. *Neural Computation* **28**:11, 2474-2504. [Abstract] [Full Text] [PDF] [PDF Plus]

1009. Benjamin Volz, Karsten Behrendt, Holger Mielenz, Igor Gilitschenski, Roland Siegwart, Juan Nieto. A data-driven approach for pedestrian intention estimation 2607-2612. [Crossref]

1010. Rui Fu, Zuo Zhang, Li Li. Using LSTM and GRU neural network methods for traffic flow prediction 324-328. [Crossref]

1011. Nicolas Papernot, Patrick McDaniel, Ananthram Swami, Richard Harang. Crafting adversarial input sequences for recurrent neural networks 49-54. [Crossref]

1012. Rui Zhao, Jinjiang Wang, Ruqiang Yan, Kezhi Mao. Machine health monitoring with LSTM networks 1-6. [Crossref]

1013. Litao Yu, Xiaoshuai Sun, Zi Huang. 2016. Robust spatial-temporal deep model for multimedia event detection. *Neurocomputing* **213**, 48-53. [Crossref]

1014. Florian Toque, Etienne Come, Mohamed Khalil El Mahrsi, Latifa Oukhellou. Forecasting dynamic public transport Origin-Destination matrices with long-Short term Memory recurrent neural networks 1071-1076. [Crossref]

1015. Mahtab Ahmed, M.A.H. Akhand, M.M. Hafizur Rahman. Handwritten Bangla Numeral Recognition Using Deep Long Short Term Memory 310-315. [Crossref]

1016. Aida Khosroshahi, Eshed Ohn-Bar, Mohan Manubhai Trivedi. Surround vehicles trajectory analysis with recurrent neural networks 2267-2272. [Crossref]

1017. Qi Liu, Tian Tan, Kai Yu. An investigation on deep learning with beta stabilizer 557-561. [Crossref]

1018. Tomah Sogabe, Haruhisa Ichikawa, Tomah Sogabe, Katsuyoshi Sakamoto, Koichi Yamaguchi, Masaru Sogabe, Takashi Sato, Yusuke Suwa. Optimization of decentralized renewable energy system by weather forecasting and deep machine learning techniques 1014-1018. [Crossref]

1019. Yang Qin, Lingfei Mo, Jing Ye, Zhening Du. Multi-channel features fitted 3D CNNs and LSTMs for human activity recognition 1-5. [Crossref]

1020. Tzu-Hsuan Yang, Tzu-Hsuan Tseng, Chia-Ping Chen. Recurrent neural network-based language models with variation in net topology, language, and granularity 71-74. [Crossref]

1021. Ming-Hsiang Su, Kun-Yi Huang, Tsung-Hsien Yang, Kuan-Jung Lai, Chung-Hsien Wu. Dialog State Tracking and action selection using deep learning mechanism for interview coaching 6-9. [Crossref]

1022. Haohao Jiang, Yao Lu, Jing Xue. Automatic Soccer Video Event Detection Based on a Deep Neural Network Combined CNN and RNN 490-494. [Crossref]

1023. Yibin Zheng, Zhengqi Wen, Bin Liu, Ya Li, Jianhua Tao. An initial research: Towards accurate pitch extraction for speech synthesis based on BLSTM 165-170. [Crossref]

1024. Min Cheng, Qian Xu, Jianming Lv, Wenyin Liu, Qing Li, Jianping Wang. MS-LSTM: A multi-scale LSTM model for BGP anomaly detection 1-6. [Crossref]

1025. Joao Canas Ferreira, Jose Fonseca. An FPGA implementation of a long short-term memory neural network 1-8. [Crossref]

1026. Shaohuai Shi, Qiang Wang, Pengfei Xu, Xiaowen Chu. Benchmarking State-of-the-Art Deep Learning Software Tools 99-104. [Crossref]

1027. Yanjie Duan, Yisheng Lv, Fei-Yue Wang. Travel time prediction with LSTM neural network 1053-1058. [Crossref]

1028. Hongwei Zhang, Feilong Bao, Guanglai Gao, Hui Zhang. Comparison on Neural Network based acoustic model in Mongolian speech recognition 1-5. [Crossref]

1029. Jarryd Son, Amit Kumar Mishra. A survey of brain inspired technologies for engineering 1-6. [Crossref]

1030. Weihua Wang, Feilong Bao, Guanglai Gao. Mongolian Named Entity Recognition with Bidirectional Recurrent Neural Networks 495-500. [Crossref]

1031. Joseph Campbell, Heni Ben Amor, Marcelo H. Ang, Georgios Fainekos. Traffic light status detection using movement patterns of vehicles 283-288. [Crossref]

1032. Kentaro Orimo, Kota Ando, Kodai Ueyoshi, Masayuki Ikebe, Tetsuya Asai, Masato Motomura. FPGA architecture for feed-forward sequential memory network targeting long-term time-series forecasting 1-6. [Crossref]

1033. Zijia Li, Chi Wun Au, Yohei Kakiuchi, Kei Okada, Masayuki Inaba. What am i doing? Robotic self-action recognition 165-170. [Crossref]

1034. Hongxin Shao, Boon-Hee Soong. Traffic flow prediction with Long Short-Term Memory Networks (LSTMs) 2986-2989. [Crossref]

1035. Yuan-yuan Chen, Yisheng Lv, Zhenjiang Li, Fei-Yue Wang. Long short-term memory model for traffic congestion prediction with online open data 132-137. [Crossref]

1036. Jing Yin, Jiancheng Lv, Yongsheng Sang, Jixiang Guo. 2016. Classification model of restricted Boltzmann machine based on reconstruction error. *Neural Computing and Applications* **2**. . [Crossref]

1037. Alex Graves, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwińska, Sergio Gómez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou, Adrià Puigdomènech Badia, Karl Moritz Hermann, Yori Zwols, Georg Ostrovski, Adam Cain, Helen King, Christopher Summerfield, Phil Blunsom, Koray Kavukcuoglu, Demis Hassabis. 2016. Hybrid computing using a neural network with dynamic external memory. *Nature* **538**:7626, 471-476. [Crossref]

1038. Lionel Pigou, Aäron van den Oord, Sander Dieleman, Mieke Van Herreweghe, Joni Dambre. 2016. Beyond Temporal Pooling: Recurrence and Temporal Convolutions for Gesture Recognition in Video. *International Journal of Computer Vision* . [Crossref]

1039. Gerhard Widmer. 2016. Getting Closer to the Essence of Music. *ACM Transactions on Intelligent Systems and Technology* **8**:2, 1-13. [Crossref]

1040. Qu Xiaoyun, Kang Xiaoning, Zhang Chao, Jiang Shuai, Ma Xiuda. Short-term prediction of wind power based on deep Long Short-Term Memory 1148-1152. [Crossref]

1041. Wenkai Dong, Ran He, Shu Zhang. Digital recognition from lip texture analysis 477-481. [Crossref]

1042. Hossein Hosseini, Sreeram Kannan, Baosen Zhang, Radha Poovendran. Learning Temporal Dependence from Time-Series Data with Latent Variables 253-262. [Crossref]

1043. Kenta Inuzuka, Tomonori Hayashi, Tomohiro Takagi. Recommendation System Based on Prediction of User Preference Changes 192-199. [Crossref]

1044. Da Zheng, Zhehuai Chen, Yue Wu, Kai Yu. Directed automatic speech transcription error correction using bidirectional LSTM 1-5. [Crossref]

1045. Tianxing He, Yu Zhang, Jasha Droppo, Kai Yu. On training bi-directional neural network language model with noise contrastive estimation 1-5. [Crossref]

1046. Runnan Li, Zhiyong Wu, Helen Meng, Lianhong Cai. DBLSTM-based multi-task learning for pitch transformation in voice conversion 1-5. [Crossref]

1047. Viet-Trung Tran, Kiem-Hieu Nguyen, Duc-Hanh Bui. A Vietnamese language model based on Recurrent Neural Network 274-278. [Crossref]

1048. Zhihe Lu, Xiang Wu, Ran He. Person identification from lip texture analysis 472-476. [Crossref]

1049. Cuong Tuan Nguyen, Masaki Nakagawa. Finite State Machine Based Decoding of Handwritten Text Using Recurrent Neural Networks 246-251. [Crossref]

1050. Brian Kenji Iwana, Volkmar Frinken, Seiichi Uchida. A Robust Dissimilarity-Based Neural Network for Temporal Pattern Recognition 265-270. [Crossref]

1051. Zhicheng Zhao, Yifan Song, Fei Su. 2016. Specific video identification via joint learning of latent semantic concept, scene and temporal structure. *Neurocomputing* **208**, 378-386. [Crossref]

1052. Paul Voigtlaender, Patrick Doetsch, Hermann Ney. Handwriting Recognition with Large Multidimensional Long Short-Term Memory Recurrent Neural Networks 228-233. [Crossref]

1053. Patrick Doetsch, Albert Zeyer, Hermann Ney. Bidirectional Decoder Networks for Attention-Based End-to-End Offline Handwriting Recognition 361-366. [Crossref]

1054. Marcus Edel, Enrico Koppe. Binarized-BLSTM-RNN based Human Activity Recognition 1-7. [Crossref]

1055. Lanhua Zhang, Huihui Yang, Chengxin Yan, Baoliang Sun. Modelling and simulation of brain memory function based on complex network and computer algorithm 1314-1318. [Crossref]

1056. Akira Sasaki, Junta Mizuno, Naoaki Okazaki, Kentaro Inui. Stance Classification by Recognizing Related Events about Targets 582-587. [Crossref]

1057. Li Sun, Tonghua Su, Ce Liu, Ruigang Wang. Deep LSTM Networks for Online Chinese Handwriting Recognition 271-276. [Crossref]

1058. Changhao Shan, Lei Xie, Kaisheng Yao. A bi-directional LSTM approach for polyphone disambiguation in Mandarin Chinese 1-5. [Crossref]

1059. Nguyen Truong Son, Le Minh Nguyen, Ho Bao Quoc, Akira Shimazu. Recognizing logical parts in legal texts using neural architectures 252-257. [Crossref]

1060. Andre Gensler, Janosch Henze, Bernhard Sick, Nils Raabe. Deep Learning for solar power forecasting — An approach using AutoEncoder and LSTM Neural Networks 002858-002865. [Crossref]

1061. Tjeng Wawan Cenggoro, Ida Siahaan. Dynamic bandwidth management based on traffic prediction using Deep Long Short Term Memory 318-323. [Crossref]

1062. AbdElRahman ElSaid, Brandon Wild, James Higgins, Travis Desell. Using LSTM recurrent neural networks to predict excess vibration events in aircraft engines 260-269. [Crossref]

1063. Yang Zhao, Tanfeng Sun, Xinghao Jiang, Shilin Wang. Long-term Residual Recurrent Network for human interaction recognition in videos 78-83. [Crossref]

1064. Daniel L. Marino, Kasun Amarasinghe, Milos Manic. Building energy load forecasting using Deep Neural Networks 7046-7051. [Crossref]

1065. Gonzalo Vaca-Castano, Samarjit Das, Joao P. Sousa, Niels D. Lobo, Mubarak Shah. 2016. Improved scene identification and object detection on egocentric vision of daily activities. *Computer Vision and Image Understanding* . [Crossref]

1066. Yuanyuan Zhao, Jie Li, Shuang Xu, Bo Xu. Investigating gated recurrent neural networks for acoustic modeling 1-5. [Crossref]

1067. Tsung-Hsien Yang, Chung-Hsien Wu, Kun-Yi Huang, Ming-Hsiang Su. Detection of mood disorder using speech emotion profiles and LSTM 1-5. [Crossref]

1068. Vidyadhar Upadhya, Avishek Chatterjee, A. P. Prathosh, Pragathi Praveena. Respiration Monitoring through Thoraco-Abdominal Video with an LSTM 165-171. [Crossref]

1069. Yujun Lu, Yihua Shi, Guimin Jia, Jinfeng Yang. A new method for semantic consistency verification of aviation radiotelephony communication based on LSTM-RNN 422-426. [Crossref]

1070. Adrian Taylor, Sylvain Leblanc, Nathalie Japkowicz. Anomaly Detection in Automobile Control Network Data with Long Short-Term Memory Networks 130-139. [Crossref]

1071. Tian Guo, Zhao Xu, Xin Yao, Haifeng Chen, Karl Aberer, Koichi Funaya. Robust Online Time Series Prediction with Recurrent Neural Networks 816-825. [Crossref]

1072. Chandranath Adak, Bidyut B. Chaudhuri, Michael Blumenstein. Offline Cursive Bengali Word Recognition Using CNNs with a Recurrent Model 429-434. [Crossref]

1073. Kerstin Unger, Laura Ackerman, Christopher H. Chatham, Dima Amso, David Badre. 2016. Working memory gating mechanisms explain developmental change in rule-guided behavior. *Cognition* **155**, 8-22. [Crossref]

1074. Qingye Ding, Zhida Li, Prerna Batta, Ljiljana Trajkovic. Detecting BGP anomalies using machine learning techniques 003352-003355. [Crossref]

1075. Dewi Suryani, Patrick Doetsch, Hermann Ney. On the Benefits of Convolutional Neural Network Combinations in Offline Handwriting Recognition 193-198. [Crossref]

1076. Mei Yuan, Yuting Wu, Li Lin. Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network 135-140. [Crossref]

1077. Jen-Tzung Chien, Alim Misbullah. Deep long short-term memory networks for speech recognition 1-5. [Crossref]

1078. Aibek Makazhanov, Zhandos Yessenbayev. Character-based feature extraction with LSTM networks for POS-tagging task 1-5. [Crossref]

1079. Kristof Varszegi. Comparison of algorithms for detecting hand movement from EEG signals 002208-002213. [Crossref]

1080. Stuner Bruno, Chatelain Clement, Paquet Thierry. A Lexicon Verification Strategy in a BLSTM Cascade Framework 234-239. [Crossref]

1081. Yujun Zhou, Bo Xu, Jiaming Xu, Lei Yang, Changliang Li, Bo Xu. Compositional Recurrent Neural Networks for Chinese Short Text Classification 137-144. [Crossref]

1082. Stuart Mason Dambrot. Exocortical Cognition: Heads in the Cloud - A transdisciplinary framework for augmenting human high-level cognitive processes 004007-004014. [Crossref]

1083. Peter U. Diehl, Guido Zarrella, Andrew Cassidy, Bruno U. Pedroni, Emre Neftci. Conversion of artificial recurrent neural networks to spiking neural networks for low-power neuromorphic hardware 1-8. [Crossref]

1084. Xueyang Wu, Su Zhu, Yue Wu, Kai Yu. Rich punctuations prediction using large-scale deep learning 1-5. [Crossref]

1085. Zenghui Sun, Lianwen Jin, Zecheng Xie, Ziyong Feng, Shuye Zhang. Convolutional Multi-directional Recurrent Network for Offline Handwritten Text Recognition 240-245. [Crossref]

1086. Anna Leontjeva, Ilya Kuzovkin. Combining Static and Dynamic Features for Multivariate Sequence Classification 21-30. [Crossref]

1087. Rui Wu, Shuli Yang, Dawei Leng, Zhenbo Luo, Yunhong Wang. Random Projected Convolutional Feature for Scene Text Recognition 132-137. [Crossref]

1088. Zhengqi Wen, Kehuang Li, Zhen Huang, Jianhua Tao, Chin-Hui Lee. Learning auxiliary categorical information for speech synthesis based on deep and recurrent neural networks 1-5. [Crossref]

1089. Wenzu Wu, Kunjin Chen, Ying Qiao, Zongxiang Lu. Probabilistic short-term wind power forecasting based on deep neural networks 1-8. [Crossref]

1090. Alfan Farizki Wicaksono, Mima Adriani. Single-output recurrent neural networks for sentence binary classification 293-296. [Crossref]

1091. Peisong Wang, Qiang Song, Hua Han, Jian Cheng. 2016. Sequentially Supervised Long Short-Term Memory for Gesture Recognition. *Cognitive Computation* **8**:5, 982-991. [Crossref]

1092. Hussein A. Al-Barazanchi, Hussam Qassim, Abhishek Verma. Novel CNN architecture with residual learning and deep supervision for large-scale scene image categorization 1-7. [Crossref]

1093. Amir Ahooye Atashin, Kamaledin Ghiasi-Shirazi, Ahad Harati. Training LDCRF model on unsegmented sequences using connectionist temporal classification 280-285. [Crossref]

1094. Daniel Paul Barrett, Ran Xu, Haonan Yu, Jeffrey Mark Siskind. 2016. Collecting and annotating the large continuous action dataset. *Machine Vision and Applications* **27**:7, 983-995. [Crossref]

1095. Minjae Lee, Kyuyeon Hwang, Jinhwan Park, Sungwook Choi, Sungho Shin, Wonyong Sung. FPGA-Based Low-Power Speech Recognition with Recurrent Neural Networks 230-235. [Crossref]

1096. Stephen Zekany, Daniel Rings, Nathan Harada, Michael A. Laurenzano, Lingjia Tang, Jason Mars. CrystalBall: Statically analyzing runtime behavior via deep sequence learning 1-12. [Crossref]

1097. Cristobal Esteban, Oliver Staeck, Stephan Baier, Yinchong Yang, Volker Tresp. Predicting Clinical Events by Combining Static and Dynamic Information Using Recurrent Neural Networks 93-101. [Crossref]

1098. Jian Kang, Wei-Qiang Zhang, Jia Liu. Lattice based transcription loss for end-to-end speech recognition 1-5. [Crossref]

1099. Alberto Bartoli, Andrea de Lorenzo, Eric Medvet, Dennis Morello, Fabiano Tarlao. "Best Dinner Ever!!!": Automatic Generation of Restaurant Reviews with LSTM-RNN 721-724. [Crossref]

1100. Kaituo Xu, Lei Xie, Kaisheng Yao. Investigating LSTM for punctuation prediction 1-5. [Crossref]

1101. Adam H. Marblestone, Greg Wayne, Konrad P. Kording. 2016. Toward an Integration of Deep Learning and Neuroscience. *Frontiers in Computational Neuroscience* **10**. . [Crossref]

1102. Xiuyi Fan, Huiguo Zhang, Cyril Leung, Chunyan Miao. Comparative study of machine learning algorithms for activity recognition with data sequence in home-like environment 168-173. [Crossref]

1103. Gang Chen, Yawei Li, Sargur N. Srihari. Word recognition with deep conditional random fields 1928-1932. [Crossref]

1104. Lamia Rahman, Nabeel Mohammed, Abul Kalam Al Azad. A new LSTM model by introducing biological cell state 1-6. [Crossref]

1105. Haimin Zhang, Min Xu. Modeling temporal information using discrete fourier transform for recognizing emotions in user-generated videos 629-633. [Crossref]

1106. Bo Yuan. Efficient hardware architecture of softmax layer in deep neural network 323-326. [Crossref]

1107. Xing Hao, Guigang Zhang, Shang Ma. 2016. Deep Learning. *International Journal of Semantic Computing* **10**:03, 417-439. [Crossref]

1108. Okko J. Rasanen, Jukka P. Saarinen. 2016. Sequence Prediction With Sparse Distributed Hyperdimensional Coding Applied to the Analysis of Mobile Phone Use Patterns. *IEEE Transactions on Neural Networks and Learning Systems* **27**:9, 1878-1889. [Crossref]

1109. Qiurui Wang, Chun Yuan. Video object segmentation by Multi-Scale Pyramidal Multi-Dimensional LSTM with generated depth context 281-285. [Crossref]

1110. Hamid Palangi, Rabab Ward, Li Deng. 2016. Distributed Compressive Sensing: A Deep Learning Approach. *IEEE Transactions on Signal Processing* **64**:17, 4504-4518. [Crossref]

1111. Minsi Wang, Li Song, Xiaokang Yang, Chuanfei Luo. A parallel-fusion RNN-LSTM architecture for image caption generation 4448-4452. [Crossref]

1112. Mohammad Daneshvar, Hadi Veisi. Persian phoneme recognition using long short-term memory neural network 111-115. [Crossref]

1113. Tao Liu, Wengang Zhou, Houqiang Li. Sign language recognition with long short-term memory 2871-2875. [Crossref]

1114. Qiaozhe Li, Xin Zhao, Kaiqi Huang. Learning temporally correlated representations using lstms for visual tracking 1614-1618. [Crossref]

1115. Eun Ju Lee, Sang Young Shin, Byoung Chul Ko, Chunho Chang. 2016. Early sinkhole detection using a drone-based thermal camera and image processing. *Infrared Physics & Technology* **78**, 223-232. [Crossref]

1116. Mohammadreza Ebrahimi, Ching Y. Suen, Olga Ormandjieva. 2016. Detecting predatory conversations in social media by deep Convolutional Neural Networks. *Digital Investigation* **18**, 33-49. [Crossref]

1117. Ram Manohar Oruganti, Shagan Sah, Suhas Pillai, Raymond Ptucha. Image description through fusion based recurrent multi-modal learning 3613-3617. [Crossref]

1118. Paul K. J. Park, Baek Hwan Cho, Jin Man Park, Kyoobin Lee, Ha Young Kim, Hyo Ah Kang, Hyun Goo Lee, Jooyeon Woo, Yohan Roh, Won Jo Lee, Chang-Woo Shin, Qiang Wang, Hyunsurk Ryu. Performance improvement of deep learning based gesture recognition using spatiotemporal demosaicing technique 1624-1628. [Crossref]

1119. Ioannis Mitliagkas, Ce Zhang, Stefan Hadjis, Christopher Re. Asynchrony begets momentum, with an application to deep learning 997-1004. [Crossref]

1120. Jiajun Cheng, Xin Zhang, Pei Li, Sheng Zhang, Zhaoyun Ding, Hui Wang. 2016. Exploring sentiment parsing of microblogging texts for opinion polling on chinese public figures. *Applied Intelligence* **45**:2, 429-442. [Crossref]

1121. Andrew Shin, Katsunori Ohnishi, Tatsuya Harada. Beyond caption to narrative: Video captioning with multiple sentences 3364-3368. [Crossref]

1122. Mike Kestemont, Guy de Pauw, Renske van Nie, Walter Daelemans. 2016. Lemmatization for variation-rich languages using deep learning. *Digital Scholarship in the Humanities* fqw034. [Crossref]

1123. Edward Choi, Andy Schuetz, Walter F Stewart, Jimeng Sun. 2016. Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association* ocw112. [Crossref]

1124. Xin Wang, Yuanchao Liu, Ming Liu, Chengjie Sun, Xiaolong Wang. 2016. Understanding Gating Operations in Recurrent Neural Networks through Opinion Expression Extraction. *Entropy* **18**:8, 294. [Crossref]

1125. Mingai Li, Meng Zhang, Xinyong Luo, Jinfu Yang. Combined long short-term memory based network employing wavelet coefficients for MI-EEG recognition 1971-1976. [Crossref]

1126. Tao Chen, Ruifeng Xu, Yulan He, Yunqing Xia, Xuan Wang. 2016. Learning User and Product Distributed Representations Using a Sequence Model for Sentiment Analysis. *IEEE Computational Intelligence Magazine* **11**:3, 34-44. [Crossref]

1127. Zhizhen Chi, Hongyang Li, Jingjing Wang, Huchuan Lu. On the importance of network architecture in training very deep neural networks 1-5. [Crossref]

1128. Paul Nickerson, Patrick Tighe, Benjamin Shickel, Parisa Rashidi. Deep neural network architectures for forecasting analgesic response 2966-2969. [Crossref]

1129. Shatha Jaradat, Nima Dokoohaki, Mihhail Matskin, Elena Ferrari. Trust and privacy correlations in social networks: A deep learning framework 203-206. [Crossref]

1130. Dong-Jin Kim, Donggeun Yoo, Bonggeun Sim, In So Kweon. Sentence learning on deep convolutional networks for image Caption Generation 246-247. [Crossref]

1131. Salvatore Vanini, Francesca Faraci, Alan Ferrari, Silvia Giordano. 2016. Using barometric pressure data to recognize vertical displacement activities on smartphones. *Computer Communications* **87**, 37-48. [Crossref]

1132. Louis Kim, Jacob Harer, Akshay Rangamani, James Moran, Philip D. Parks, Alik Widge, Emad Eskandar, Darin Dougherty, Sang Peter Chin. Predicting local field potentials with recurrent neural networks 808-811. [Crossref]

1133. Santiago Pascual, Antonio Bonafonte. Multi-output RNN-LSTM for multiple speaker speech synthesis and adaptation 2325-2329. [Crossref]

1134. Shuohao Li, Jun Zhang, Qiang Guo, Jun Lei, Dan Tu. Generating video description with Long-Short Term Memory 73-78. [Crossref]

1135. Catherine Inibhunu, Scott Langevin, Scott Ralph, Nathan Kronefeld, Harold Soh, Greg A. Jamieson, Scott Sanner, Sean W. Kortschot, Chelsea Carrasco, Madeleine White. Adapting level of detail in user interfaces for Cybersecurity operations 13-16. [Crossref]

1136. Shaun K. Ryman, Neil D.B. Bruce, Michael S. Freund. 2016. Temporal responses of chemically diverse sensor arrays for machine olfaction using artificial intelligence. *Sensors and Actuators B: Chemical* **231**, 666-674. [Crossref]

1137. Seonwoo Min, Byunghan Lee, Sungroh Yoon. 2016. Deep learning in bioinformatics. *Briefings in Bioinformatics* bbw068. [Crossref]

1138. A. K. Deklel, A. M. Hamdy, E. M. Saad. 2016. Multi-objective symbolic regression using long-term artificial neural network memory (LTANN-MEM) and neural symbolization algorithm (NSA). *Neural Computing and Applications* . [Crossref]

1139. F. Fernández-Navarro, Maria Angeles de la Cruz, P. A. Gutiérrez, A. Castaño, C. Hervás-Martínez. 2016. Time series forecasting by recurrent product unit neural networks. *Neural Computing and Applications* . [Crossref]

1140. Boya Wu, Jia Jia, Tao He, Juan Du, Xiaoyuan Yi, Yishuang Ning. Inferring users' emotions for human-mobile voice dialogue applications 1-6. [Crossref]

1141. Peerapon Vateekul, Thanabhat Koomsubha. A study of sentiment analysis using deep learning techniques on Thai Twitter data 1-6. [Crossref]

1142. Shuo Hong Wang, Xi En Cheng, Yan Qiu Chen. Tracking undulatory body motion of multiple fish based on midline dynamics modeling 1-6. [Crossref]

1143. Dharshan Kumaran, Demis Hassabis, James L. McClelland. 2016. What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences* **20**:7, 512-534. [Crossref]

1144. Chiori Hori, Shinji Watanabe, Takaaki Hori, Bret A. Harsham, JohnR. Hershey, Yusuke Koji, Yoichi Fujii, Yuki Furumoto. Driver confusion status detection using recurrent neural networks 1-6. [Crossref]

1145. Wonyong Sung, Jinhwan Park. Architecture exploration of a programmable neural network processor for embedded systems 124-131. [Crossref]

1146. Jingyi Hou, Xinxiao Wu, Feiwu Yu, Yunde Jia. Multimedia event detection via deep spatial-temporal neural networks 1-6. [Crossref]

1147. Xi En Cheng, Shuo Hong Wang, Yan Qiu Chen. 3D tracking targets via kinematic model weighted particle filter 1-6. [Crossref]

1148. Xudong Li, Mao Ye, Dan Liu, Feng Zhang, Song Tang. Memory-based object detection in surveillance scenes 1-6. [Crossref]

1149. Ju-Youn Park, Yong-Ho Yoo, Deok-Hwa Kim, Jong-Hwan Kim. Integrated adaptive resonance theory neural model for episodic memory with task memory for task performance of robots 4873-4880. [Crossref]

1150. Haijun Zhang, Jingxuan Li, Yuzhu Ji, Heng Yue. A character-level sequence-to-sequence method for subtitle learning 780-783. [Crossref]

1151. Angelica I. Aviles, Samar M. Alsaleh, Eduard Montseny, Pilar Sobrevilla, Alicia Casals. A Deep-Neuro-Fuzzy approach for estimating the interaction forces in Robotic surgery 1113-1119. [Crossref]

1152. Qiuwen Chen, Qinru Qiu. Enhancing bidirectional association between deep image representations and loosely correlated texts 3164-3171. [Crossref]

1153. Qing-Yu Zhou, De-Quan Zheng, Tie-Jun Zhao. SLDP: Sequence learning dependency parsing model using long short-term memory 111-116. [Crossref]

1154. Zhenqi Xu, Jiani Hu, Weihong Deng. Recurrent convolutional neural network for video classification 1-6. [Crossref]

1155. Xinhui Song, Ke Chen, Jie Lei, Li Sun, Zhiyuan Wang, Lei Xie, Mingli Song. Category driven deep recurrent neural network for video summarization 1-6. [Crossref]

1156. Saurav Gupta, Nitin Anand Shrivastava, Abbas Khosravi, Bijaya Ketan Panigrahi. Wind ramp event prediction with parallelized gradient boosted regression trees 5296-5301. [Crossref]

1157. SungMahn Ahn. 2016. Deep Learning Architectures and Applications. *Journal of Intelligence and Information Systems* **22**:2, 127-142. [Crossref]

1158. Martin V. Butz. 2016. Toward a Unified Sub-symbolic Computational Theory of Cognition. *Frontiers in Psychology* **7**. . [Crossref]

1159. Wei-Xin Li, Nuno Vasconcelos. 2016. Complex Activity Recognition Via Attribute Dynamics. *International Journal of Computer Vision* **43**. . [Crossref]

1160. Leonard Truong, Rajkishore Barik, Ehsan Totoni, Hai Liu, Chick Markley, Armando Fox, Tatiana Shpeisman. 2016. Latte: a language, compiler, and runtime for elegant and efficient deep neural networks. *ACM SIGPLAN Notices* **51**:6, 209-223. [Crossref]

1161. Haonan Yu, Jiang Wang, Zhiheng Huang, Yi Yang, Wei Xu. Video Paragraph Captioning Using Hierarchical Recurrent Neural Networks 4584-4593. [Crossref]

1162. Stefanos Zafeiriou, Athanasios Papaioannou, Irene Kotsia, Mihalis Nicolaou, Guoying Zhao. Facial Affect "In-the-Wild": A Survey and a New Database 1487-1498. [Crossref]

1163. Jiang Wang, Yi Yang, Junhua Mao, Zhiheng Huang, Chang Huang, Wei Xu. CNN-RNN: A Unified Framework for Multi-label Image Classification 2285-2294. [Crossref]

1164. Zeynep Akata, Mateusz Malinowski, Mario Fritz, Bernt Schiele. Multi-cue Zero-Shot Learning with Strong Supervision 59-68. [Crossref]

1165. Jun Yuan, Bingbing Ni, Xiaokang Yang, Ashraf A. Kassim. Temporal Action Localization with Pyramid of Score Distribution Features 3093-3102. [Crossref]

1166. Yuncheng Li, Yale Song, Liangliang Cao, Joel Tetreault, Larry Goldberg, Alejandro Jaimes, Jiebo Luo. TGIF: A New Dataset and Benchmark on Animated GIF Description 4641-4650. [Crossref]

1167. Qi Wu, Peng Wang, Chunhua Shen, Anthony Dick, Anton van den Hengel. Ask Me Anything: Free-Form Visual Question Answering Based on Knowledge from External Sources 4622-4630. [Crossref]

1168. Chen-Yu Lee, Simon Osindero. Recursive Recurrent Nets with Attention Modeling for OCR in the Wild 2231-2239. [Crossref]

1169. Xiaodan Liang, Xiaohui Shen, Donglai Xiang, Jiashi Feng, Liang Lin, Shuicheng Yan. Semantic Object Parsing with Local-Global Long Short-Term Memory 3185-3193. [Crossref]

1170. Guo-Jun Qi. Hierarchically Gated Deep Networks for Semantic Segmentation 2267-2275. [Crossref]

1171. Baoguang Shi, Xinggang Wang, Pengyuan Lyu, Cong Yao, Xiang Bai. Robust Scene Text Recognition with Automatic Rectification 4168-4176. [Crossref]

1172. Hexiang Hu, Guang-Tong Zhou, Zhiwei Deng, Zicheng Liao, Greg Mori. Learning Structured Inference Neural Networks with Label Relations 2960-2968. [Crossref]

1173. Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, Silvio Savarese. Social LSTM: Human Trajectory Prediction in Crowded Spaces 961-971. [Crossref]

1174. Song Han, Xingyu Liu, Huizi Mao, Jing Pu, Ardavan Pedram, Mark A. Horowitz, William J. Dally. EIE: Efficient Inference Engine on Compressed Deep Neural Network 243-254. [Crossref]

1175. Hoo-Chang Shin, Kirk Roberts, Le Lu, Dina Demner-Fushman, Jianhua Yao, Ronald M Summers. Learning to Read Chest X-Rays: Recurrent Neural Cascade Model for Automated Image Annotation 2497-2506. [Crossref]

1176. Russell Stewart, Mykhaylo Andriluka, Andrew Y. Ng. End-to-End People Detection in Crowded Scenes 2325-2333. [Crossref]

1177. Bharat Singh, Tim K. Marks, Michael Jones, Oncel Tuzel, Ming Shao. A Multi-stream Bi-directional Recurrent Neural Network for Fine-Grained Action Detection 1961-1970. [Crossref]

1178. Alessandro Bay, Skjalg Lepsoy, Enrico Magli. Stable limit cycles in recurrent neural networks 89-92. [Crossref]

1179. Pablo Torres, Carlos Catania, Sebastian Garcia, Carlos Garcia Garino. An analysis of Recurrent Neural Networks for Botnet detection behavior 1-6. [Crossref]

1180. Chao An, Jiuming Huang, Shoufeng Chang, Zhijie Huang. Question Similarity Modeling with Bidirectional Long Short-Term Memory Neural Network 318-322. [Crossref]

1181. Sean Bell, C. Lawrence Zitnick, Kavita Bala, Ross Girshick. Inside-Outside Net: Detecting Objects in Context with Skip Pooling and Recurrent Neural Networks 2874-2883. [Crossref]

1182. Yuanping Nie, Chao An, Jiuming Huang, Zhou Yan, Yi Han. A Bidirectional LSTM Model for Question Title and Body Analysis in Question Answering 307-311. [Crossref]

1183. Andrew Owens, Phillip Isola, Josh McDermott, Antonio Torralba, Edward H. Adelson, William T. Freeman. Visually Indicated Sounds 2405-2413. [Crossref]

1184. Yuke Zhu, Oliver Groth, Michael Bernstein, Li Fei-Fei. Visual7W: Grounded Question Answering in Images 4995-5004. [Crossref]

1185. Wei Wang, Zhen Cui, Yan Yan, Jiashi Feng, Shuicheng Yan, Xiangbo Shu, Nicu Sebe. Recurrent Face Aging 2378-2386. [Crossref]

1186. Jacob Andreas, Marcus Rohrbach, Trevor Darrell, Dan Klein. Neural Module Networks 39-48. [Crossref]

1187. Yingwei Li, Weixin Li, Vijay Mahadevan, Nuno Vasconcelos. VLAD3: Encoding Dynamics of Deep Features for Action Recognition 1951-1960. [Crossref]

1188. Qi Wu, Chunhua Shen, Lingqiao Liu, Anthony Dick, Anton van den Hengel. What Value Do Explicit High Level Concepts Have in Vision to Language Problems? 203-212. [Crossref]

1189. Mostafa S. Ibrahim, Srikanth Muralidharan, Zhiwei Deng, Arash Vahdat, Greg Mori. A Hierarchical Deep Temporal Model for Group Activity Recognition 1971-1980. [Crossref]

1190. Jiajun Cheng, Pei Li, Zhaoyun Ding, Sheng Zhang, Hui Wang. Sentiment Classification of Chinese Microblogging Texts with Global RNN 653-657. [Crossref]

1191. Zhen Cui, Shengtao Xiao, Jiashi Feng, Shuicheng Yan. Recurrently Target-Attending Tracking 1449-1458. [Crossref]

1192. Zichao Yang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Smola. Stacked Attention Networks for Image Question Answering 21-29. [Crossref]

1193. Basura Fernando, Peter Anderson, Marcus Hutter, Stephen Gould. Discriminative Hierarchical Rank Pooling for Activity Recognition 1924-1932. [Crossref]

1194. Scott Reed, Zeynep Akata, Honglak Lee, Bernt Schiele. Learning Deep Representations of Fine-Grained Visual Descriptions 49-58. [Crossref]

1195. Amir Shahroudy, Jun Liu, Tian-Tsong Ng, Gang Wang. NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis 1010-1019. [Crossref]

1196. Hyeonwoo Noh, Paul Hongsuck Seo, Bohyung Han. Image Question Answering Using Convolutional Neural Network with Dynamic Parameter Prediction 30-38. [Crossref]

1197. Albert Haque, Alexandre Alahi, Li Fei-Fei. Recurrent Attention Models for Depth-Based Person Identification 1229-1238. [Crossref]

1198. Bingbing Ni, Xiaokang Yang, Shenghua Gao. Progressively Parsing Interactional Objects for Fine Grained Action Detection 1020-1028. [Crossref]

1199. Guo-Bing Zhou, Jianxin Wu, Chen-Lin Zhang, Zhi-Hua Zhou. 2016. Minimal gated unit for recurrent neural networks. *International Journal of Automation and Computing* **13**:3, 226-234. [Crossref]

1200. Lisa Anne Hendricks, Subhashini Venugopalan, Marcus Rohrbach, Raymond Mooney, Kate Saenko, Trevor Darrell. Deep Compositional Captioning: Describing Novel Object Categories without Paired Training Data 1-10. [Crossref]

1201. Egon Kidmose, Matija Stevanovic, Jens Myrup Pedersen. Correlating intrusion detection alerts on bot malware infections using neural network 1-8. [Crossref]

1202. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition 770-778. [Crossref]

1203. Ryo Akita, Akira Yoshihara, Takashi Matsubara, Kuniaki Uehara. Deep learning for stock prediction using numerical and textual information 1-6. [Crossref]

1204. Ishan Misra, C. Lawrence Zitnick, Margaret Mitchell, Ross Girshick. Seeing through the Human Reporting Bias: Visual Classifiers from Noisy Human-Centric Labels 2930-2939. [Crossref]

1205. Yuancheng Ye, Yingli Tian. Embedding Sequential Information into Spatiotemporal Features for Action Recognition 1110-1118. [Crossref]

1206. Yingwei Pan, Tao Mei, Ting Yao, Houqiang Li, Yong Rui. Jointly Modeling Embedding and Translation to Bridge Video and Language 4594-4602. [Crossref]

1207. Chuang Gan, Ting Yao, Kuiyuan Yang, Yi Yang, Tao Mei. You Lead, We Exceed: Labor-Free Video Concept Learning by Jointly Exploiting Web Videos and Images 923-932. [Crossref]

1208. Yuusuke Kataoka, Takashi Matsubara, Kuniaki Uehara. Image generation using generative adversarial networks and attention mechanism 1-6. [Crossref]

1209. Hakan Bilen, Basura Fernando, Efstratios Gavves, Andrea Vedaldi, Stephen Gould. Dynamic Image Networks for Action Recognition 3034-3042. [Crossref]

1210. Justin Johnson, Andrej Karpathy, Li Fei-Fei. DenseCap: Fully Convolutional Localization Networks for Dense Captioning 4565-4574. [Crossref]

1211. Wang Yue, Wang Xiaojie, Mao Yuzhao. 2016. First-Feed LSTM model for video description. *The Journal of China Universities of Posts and Telecommunications* **23**:3, 89-93. [Crossref]

1212. Liang-Chieh Chen, Jonathan T. Barron, George Papandreou, Kevin Murphy, Alan L. Yuille. Semantic Image Segmentation with Task-Specific Edge Detection Using CNNs and a Discriminatively Trained Domain Transform 4545-4554. [Crossref]

1213. Behrooz Mahasseni, Sinisa Todorovic. Regularizing Long Short Term Memory with 3D Human-Skeleton Sequences for Action Recognition 3054-3062. [Crossref]

1214. Pingbo Pan, Zhongwen Xu, Yi Yang, Fei Wu, Yueting Zhuang. Hierarchical Recurrent Neural Encoder for Video Representation with Application to Captioning 1029-1038. [Crossref]

1215. Zae Kim, Young-Seob Jeong, Hyung Oh, Kyo-Joong Oh, Chae-Gyun Lim, Youssef Iraqi, Ho-Jin Choi. 2016. Investigating the Impact of Possession-Way of a Smartphone on Action Recognition. *Sensors* **16**:6, 812. [Crossref]

1216. Sebastian Lapuschkin, Alexander Binder, Gregoire Montavon, Klaus-Robert Muller, Wojciech Samek. Analyzing Classifiers: Fisher Vectors and Deep Neural Networks 2912-2920. [Crossref]

1217. Ronghang Hu, Huazhe Xu, Marcus Rohrbach, Jiashi Feng, Kate Saenko, Trevor Darrell. Natural Language Object Retrieval 4555-4564. [Crossref]

1218. Yang-hoon Kim, Yong-keun Hwang, Tae-gwan Kang, Kyo-min Jung. 2016. LSTM Language Model Based Korean Sentence Generation. *The Journal of Korean Institute of Communications and Information Sciences* **41**:5, 592-601. [Crossref]

1219. Yackov Lubarsky, Amit Wolf, Lior Wolf, Curime Batliner, Jake Newsum. Softbot: Software-based lead-through for rigid servo robots 2401-2407. [Crossref]

1220. Bo Fan, Lei Xie, Shan Yang, Lijuan Wang, Frank K. Soong. 2016. A deep bidirectional LSTM approach for video-realistic talking head. *Multimedia Tools and Applications* **75**:9, 5287-5309. [Crossref]

1221. Ashesh Jain, Avi Singh, Hema S Koppula, Shane Soh, Ashutosh Saxena. Recurrent Neural Networks for driver activity anticipation via sensory-fusion architecture 3118-3125. [Crossref]

1222. Hayit Greenspan, Bram van Ginneken, Ronald M. Summers. 2016. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Transactions on Medical Imaging* **35**:5, 1153-1159. [Crossref]

1223. Michael S. Gashler, Stephen C. Ashmore. 2016. Modeling time series data with deep Fourier neural networks. *Neurocomputing* **188**, 3-11. [Crossref]

1224. Marvin Zhang, Zoe McCarthy, Chelsea Finn, Sergey Levine, Pieter Abbeel. Learning deep neural network policies with continuous memory states 520-527. [Crossref]

1225. Yang Gao, Lisa Anne Hendricks, Katherine J. Kuchenbecker, Trevor Darrell. Deep learning for tactile understanding from visual and haptic data 536-543. [Crossref]

1226. Ricard Marxer, Hendrik Purwins. 2016. Unsupervised Incremental Online Learning and Prediction of Musical Audio Signals. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **24**:5, 863-874. [Crossref]

1227. Victor C. Liang, Richard T. B. Ma, Wee Siong Ng, Li Wang, Marianne Winslett, Huayu Wu, Shanshan Ying, Zhenjie Zhang. Mercury: Metro density prediction with recurrent neural network on streaming CDR data 1374-1377. [Crossref]

1228. Sebastian Otte, Christian Weiss, Tobias Scherer, Andreas Zell. Recurrent Neural Networks for fast and robust vibration-based ground classification on mobile robots 5603-5608. [Crossref]

1229. Sheridan S. Curley, Richard E. Harang. Grammatical Inference and Machine Learning Approaches to Post-Hoc LangSec 171-178. [Crossref]

1230. Xu Xu, Hao Ge, Shenghong Li. An improvement on recurrent neural network by combining convolution neural network and a simple initialization of the weights 150-154. [Crossref]

1231. Costas Sideris, Haik Kalantarian, Ebrahim Nemati, Majid Sarrafzadeh. Building Continuous Arterial Blood Pressure Prediction Models Using Recurrent Networks 1-5. [Crossref]

1232. Kobus Barnard. 2016. Computational Methods for Integrating Vision and Language. *Synthesis Lectures on Computer Vision* **6**:1, 1-227. [Crossref]

1233. Yohannes Kassahun, Bingbin Yu, Abraham Temesgen Tibebu, Danail Stoyanov, Stamatia Giannarou, Jan Hendrik Metzen, Emmanuel Vander Poorten. 2016. Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions. *International Journal of Computer Assisted Radiology and Surgery* **11**:4, 553-568. [Crossref]

1234. Anvardh Nanduri, Lance Sherry. Anomaly detection in aircraft data using Recurrent Neural Networks (RNN) 5C2-1-5C2-8. [Crossref]

1235. Zhong-Qiu Wang, DeLiang Wang. 2016. A Joint Training Framework for Robust Automatic Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **24**:4, 796-806. [Crossref]

1236. Pablo Peso Parada, Dushyant Sharma, Jose Lainez, Daniel Barreda, Toon van Waterschoot, Patrick A. Naylor. 2016. A Single-Channel Non-Intrusive C50 Estimator Correlated With Speech Recognition Performance. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **24**:4, 719-732. [Crossref]

1237. Hamid Palangi, Li Deng, Yelong Shen, Jianfeng Gao, Xiaodong He, Jianshu Chen, Xinying Song, Rabab Ward. 2016. Deep Sentence Embedding Using Long Short-Term Memory Networks: Analysis and Application to Information Retrieval. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **24**:4, 694-707. [Crossref]

1238. Saad Bin Ahmed, Saeeda Naz, Muhammad Imran Razzak, Shiekh Faisal Rashid, Muhammad Zeeshan Afzal, Thomas M. Breuel. 2016. Evaluation of cursive and non-cursive scripts using recurrent neural networks. *Neural Computing and Applications* **27**:3, 603-613. [Crossref]

1239. Minesh Mathew, Ajeet Kumar Singh, C. V. Jawahar. Multilingual OCR for Indic Scripts 186-191. [Crossref]

1240. Fallak Asad, Adnan Ul-Hasan, Faisal Shafait, Andreas Dengel. High Performance OCR for Camera-Captured Blurred Documents with LSTM Networks 7-12. [Crossref]

1241. A. L. Edelen, S. G. Biedron, B. E. Chase, D. Edstrom, S. V. Milton, P. Stabile. 2016. Neural Networks for Modeling and Control of Particle Accelerators. *IEEE Transactions on Nuclear Science* **63**:2, 878-897. [Crossref]

1242. Martin Schall, Marc-Peter Schambach, Matthias O. Franz. Increasing Robustness of Handwriting Recognition Using Character N-Gram Decoding on Large Lexica 156-161. [Crossref]

1243. Jeff Hawkins, Subutai Ahmad. 2016. Why Neurons Have Thousands of Synapses, a Theory of Sequence Memory in Neocortex. *Frontiers in Neural Circuits* **10**. . [Crossref]

1244. Jaeju Kim, Hwansoo Han. 2016. Neural Predictive Coding for Text Compression Using GPGPU. *KIISE Transactions on Computing Practices* **22**:3, 127-132. [Crossref]

1245. Michael Maynord, Sambit Bhattacharya, David W. Aha. Image Surveillance Assistant 1-7. [Crossref]

1246. Xinxing Li, Haishu Xianyu, Jiashen Tian, Wenxiao Chen, Fanhang Meng, Mingxing Xu, Lianhong Cai. A deep bidirectional long short-term memory based multi-scale approach for music dynamic emotion prediction 544-548. [Crossref]

1247. Jinyu Li, Abdelrahman Mohamed, Geoffrey Zweig, Yifan Gong. Exploring multidimensional lstms for large vocabulary ASR 4940-4944. [Crossref]

1248. Zhizheng Wu, Simon King. Investigating gated recurrent networks for speech synthesis 5140-5144. [Crossref]

1249. Cesar Laurent, Gabriel Pereyra, Philemon Brakel, Ying Zhang, Yoshua Bengio. Batch normalized recurrent neural networks 2657-2661. [Crossref]

1250. Zhiying Huang, Jian Tang, Shaofei Xue, Lirong Dai. Speaker adaptation OF RNN-BLSTM for speech recognition based on speaker code 5305-5309. [Crossref]

1251. George Trigeorgis, Fabien Ringeval, Raymond Brueckner, Erik Marchi, Mihalis A. Nicolaou, Bjorn Schuller, Stefanos Zafeiriou. Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network 5200-5204. [Crossref]

1252. Yajie Miao, Mohammad Gowayyed, Xingyu Na, Tom Ko, Florian Metze, Alexander Waibel. An empirical exploration of CTC acoustic models 2623-2627. [Crossref]

1253. Tianxing He, Jasha Droppo. Exploiting LSTM structure in deep neural networks for speech recognition 5445-5449. [Crossref]

1254. Xinyu Lan, Xu Li, Yishuang Ning, Zhiyong Wu, Helen Meng, Jia Jia, Lianhong Cai. Low level descriptors based DBLSTM bottleneck feature for speech driven talking avatar 5550-5554. [Crossref]

1255. Jonathan A. Tepper, Mahmud S. Shertil, Heather M. Powell. 2016. On the importance of sluggish state memory for learning long term dependency. *Knowledge-Based Systems* **96**, 104-114. [Crossref]

1256. Kyuyeon Hwang, Wonyong Sung. Character-level incremental speech recognition with recurrent neural networks 5335-5339. [Crossref]

1257. Michael Wand, Jan Koutnik, Jurgen Schmidhuber. Lipreading with long short-term memory 6115-6119. [Crossref]

1258. Dapeng Tao, Xu Lin, Lianwen Jin, Xuelong Li. 2016. Principal Component 2-D Long Short-Term Memory for Font Recognition on Single Chinese Characters. *IEEE Transactions on Cybernetics* **46**:3, 756-765. [Crossref]

1259. Truyen Van Phan, Masaki Nakagawa. 2016. Combination of global and local contexts for text/non-text classification in heterogeneous online handwritten documents. *Pattern Recognition* **51**, 112-124. [Crossref]

1260. Wenfu Wang, Shuang Xu, Bo Xu. Gating recurrent mixture density networks for acoustic modeling in statistical parametric speech synthesis 5520-5524. [Crossref]

1261. Colin Raffel, Daniel P. W. Ellis. Pruning subsequence search with attention-based embedding 554-558. [Crossref]

1262. Zhiyuan Tang, Dong Wang, Zhiyong Zhang. Recurrent neural network training with dark knowledge transfer 5900-5904. [Crossref]

1263. Shubhankar Potdar, Abhishek Sawarkar, Faruk Kazi. Learning by demonstration from multiple agents in humanoid robots 1-6. [Crossref]

1264. Suman Ravuri, Andreas Stolcke. A comparative study of recurrent neural network models for lexical domain classification 6075-6079. [Crossref]

1265. Georg Heigold, Ignacio Moreno, Samy Bengio, Noam Shazeer. End-to-end text-dependent speaker verification 5115-5119. [Crossref]

1266. Eshed Ohn-Bar, Mohan Manubhai Trivedi. 2016. Looking at Humans in the Age of Self-Driving and Highly Automated Vehicles. *IEEE Transactions on Intelligent Vehicles* **1**:1, 90-104. [Crossref]

1267. Chaojun Liu, Yongqiang Wang, Kshitiz Kumar, Yifan Gong. Investigations on speaker adaptation of LSTM RNN models for speech recognition 5020-5024. [Crossref]

1268. Shashank Jaiswal, Michel Valstar. Deep learning the dynamic appearance and shape of facial action units 1-8. [Crossref]

1269. Yishan Jiao, Ming Tu, Visar Berisha, Julie Liss. Online speaking rate estimation using recurrent neural networks 5245-5249. [Crossref]

1270. Takaaki Hori, Chiori Hori, Shinji Watanabe, John R. Hershey. Minimum word error training of long short-term memory recurrent neural network language models for speech recognition 5990-5994. [Crossref]

1271. William Chan, Navdeep Jaitly, Quoc Le, Oriol Vinyals. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition 4960-4964. [Crossref]

1272. Sungho Shin, Kyuyeon Hwang, Wonyong Sung. Fixed-point performance analysis of recurrent neural networks 976-980. [Crossref]

1273. Yun Wang, Leonardo Neves, Florian Metze. Audio-based multimedia event detection using deep recurrent neural networks 2742-2746. [Crossref]

1274. Tian Tan, Yanmin Qian, Dong Yu, Souvik Kundu, Liang Lu, Khe Chai Sim, Xiong Xiao, Yu Zhang. Speaker-aware training of LSTM-RNNS for acoustic modelling 5280-5284. [Crossref]

1275. David Haws, Dimitrios Dimitriadis, George Saon, Samuel Thomas, Michael Picheny. On the importance of event detection for ASR 5705-5709. [Crossref]

1276. Zhiyun Lu, Vikas Sindhwani, Tara N. Sainath. Learning compact recurrent neural networks 5960-5964. [Crossref]

1277. Mandy Korpusik, Calvin Huang, Michael Price, James Glass. Distributional semantics for understanding spoken meal descriptions 6070-6074. [Crossref]

1278. Yaodong Tang, Yuchen Huang, Zhiyong Wu, Helen Meng, Mingxing Xu, Lianhong Cai. Question detection from acoustic features using recurrent neural network with gated recurrent unit 6125-6129. [Crossref]

1279. Haishu Xianyu, Xinxing Li, Wenxiao Chen, Fanhang Meng, Jiashen Tian, Mingxing Xu, Lianhong Cai. SVR based double-scale regression for dynamic emotion prediction in music 549-553. [Crossref]

1280. Sibo Tong, Hao Gu, Kai Yu. A comparative study of robustness of deep learning approaches for VAD 5695-5699. [Crossref]

1281. Yu Zhang, Guoguo Chen, Dong Yu, Kaisheng Yaco, Sanjeev Khudanpur, James Glass. Highway long short-term memory RNNS for distant speech recognition 5755-5759. [Crossref]

1282. Quanjie Yu, Peng Liu, Zhiyong Wu, Shiyin K Ang, Helen Meng, Lianhong Cai. Learning cross-lingual information with multilingual BLSTM for speech synthesis of low-resource languages 5545-5549. [Crossref]

1283. Shi-Xiong Zhang, Rui Zhao, Chaojun Liu, Jinyu Li, Yifan Gong. Recurrent support vector machines for speech recognition 5885-5889. [Crossref]

1284. Eunbyung Park, Xufeng Han, Tamara L. Berg, Alexander C. Berg. Combining multiple sources of knowledge in deep CNNs for action recognition 1-8. [Crossref]

1285. H. Francis Song, Guangyu R. Yang, Xiao-Jing Wang. 2016. Training Excitatory-Inhibitory Recurrent Neural Networks for Cognitive Tasks: A Simple and Flexible Framework. *PLOS Computational Biology* **12**:2, e1004792. [Crossref]

1286. Saeeda Naz, Arif I. Umar, Riaz Ahmad, Saad B. Ahmed, Syed H. Shirazi, Imran Siddiqi, Muhammad I. Razzak. 2016. Offline cursive Urdu-Nastaliq script recognition using multidimensional recurrent neural networks. *Neurocomputing* **177**, 228-241. [Crossref]

1287. Hans Krupakar, Keerthika Rajvel, B Bharathi, S Angel Deborah, Vallidevi Krishnamurthy. A survey of voice translation methodologies — Acoustic dialect decoder 1-9. [Crossref]

1288. Vassilis Vassiliades, Chris Christodoulou. 2016. Behavioral plasticity through the modulation of switch neurons. *Neural Networks* **74**, 35-51. [Crossref]

1289. Roman Melnyk, Arsenii Zawyalow. Image retrieval by statistical features and artificial neural networks 706-708. [Crossref]

1290. Suraj Srinivas, Ravi Kiran Sarvadevabhatla, Konda Reddy Mopuri, Nikita Prabhu, Srinivas S. S. Kruthiventi, R. Venkatesh Babu. 2016. A Taxonomy of Deep Convolutional Neural Nets for Computer Vision. *Frontiers in Robotics and AI* **2**. . [Crossref]

1291. J. Ignacio Toledo, Sebastian Sudholt, Alicia Fornés, Jordi Cucurull, Gernot A. Fink, Josep Lladós. Handwritten Word Image Categorization with Convolutional Neural Networks and Spatial Pyramid Pooling 543-552. [Crossref]

1292. Kun Li, Yumei Chai, Hongling Zhao, Xiaofei Nan, Yueshu Zhao. Learning to Recognize Protected Health Information in Electronic Health Records with Recurrent Neural Network 575-582. [Crossref]

1293. Ngoc Khuong Nguyen, Anh-Cuong Le, Hong Thai Pham. Deep Bi-directional Long Short-Term Memory Neural Networks for Sentiment Analysis of Social Data 255-268. [Crossref]

1294. Aniruddha Kembhavi, Mike Salvato, Eric Kolve, Minjoon Seo, Hannaneh Hajishirzi, Ali Farhadi. A Diagram is Worth a Dozen Images 235-251. [Crossref]

1295. Ralf Schlüter, Patrick Doetsch, Pavel Golik, Markus Kitza, Tobias Menne, Kazuki Irie, Zoltán Tüske, Albert Zeyer. Automatic Speech Recognition Based on Neural Networks 3-17. [Crossref]

1296. Erik Gawehn, Jan A. Hiss, Gisbert Schneider. 2016. Deep Learning in Drug Discovery. *Molecular Informatics* **35**:1, 3-14. [Crossref]

1297. Zhiwen Xie, Zhao Zeng, Guangyou Zhou, Tingting He. Knowledge Base Question Answering Based on Deep Learning Models 300-311. [Crossref]

1298. Santiago Pascual, Antonio Bonafonte. Prosodic Break Prediction with RNNs 64-72. [Crossref]

1299. Florian Eyben. Acoustic Features and Modelling 9-122. [Crossref]

1300. Chuanhai Dong, Jiajun Zhang, Chengqing Zong, Masanori Hattori, Hui Di. Character-Based LSTM-CRF with Radical-Level Features for Chinese Named Entity Recognition 239-250. [Crossref]

1301. Trang Pham, Truyen Tran, Dinh Phung, Svetha Venkatesh. Faster training of very deep networks via p-norm gates 3542-3547. [Crossref]

1302. Qikang Wei, Tao Chen, Ruifeng Xu, Yulan He, Lin Gui. 2016. Disease named entity recognition by combining conditional random fields and bidirectional recurrent neural networks. *Database* **2016**, baw140. [Crossref]

1303. Peter P. Groumpos. 2016. Deep Learning vs. Wise Learning: A Critical and Challenging Overview. *IFAC-PapersOnLine* **49**:29, 180-189. [Crossref]

1304. Zuo-wei Wang. 2016. A Self-Organizing Incremental Spatiotemporal Associative Memory Networks Model for Problems with Hidden State. *Computational Intelligence and Neuroscience* **2016**, 1-14. [Crossref]

1305. Yi Yang, Jingyun Zhang, Jiasong Sun. Speech Activity Detection and Speaker Localization Based on Distributed Microphones 392-400. [Crossref]

1306. Rahul Rama Varior, Bing Shuai, Jiwen Lu, Dong Xu, Gang Wang. A Siamese Long Short-Term Memory Architecture for Human Re-identification 135-153. [Crossref]

1307. Jieru Mei, Luo Dai, Baoguang Shi, Xiang Bai. Scene text script identification with Convolutional Recurrent Neural Networks 4053-4058. [Crossref]

1308. Jiani Zhang, Irwin King. Topological Order Discovery via Deep Knowledge Tracing 112-119. [Crossref]

1309. Yangyang Lu, Ge Li, Rui Miao, Zhi Jin. Learning Embeddings of API Tokens to Facilitate Deep Learning Based Program Processing 527-539. [Crossref]

1310. Guy Lev, Gil Sadeh, Benjamin Klein, Lior Wolf. **9910**, 833. [Crossref]

1311. Mujiono Sadikin, Mohamad Ivan Fanany, T. Basaruddin. 2016. A New Data Representation Based on Training Data Characteristics to Extract Drug Name Entity in Medical Text. *Computational Intelligence and Neuroscience* **2016**, 1-16. [Crossref]

1312. Yipin Zhou, Tamara L. Berg. Learning Temporal Transformations from Time-Lapse Videos 262-277. [Crossref]

1313. Maximilian Panzner, Philipp Cimiano. Comparing Hidden Markov Models and Long Short Term Memory Neural Networks for Learning Action Representations 94-105. [Crossref]

1314. Tonghua Su, Shukai Jia, Qiufeng Wang, Li Sun, Ruigang Wang. Novel character segmentation method for overlapped Chinese handwriting recognition based on LSTM neural networks 1141-1146. [Crossref]

1315. Bo Sun, Qihua Xu, Jun He, Lejun Yu, Liandong Li, Qinglan Wei. Audio-Video Based Multimodal Emotion Recognition Using SVMs and Deep Learning 621-631. [Crossref]

1316. Jun Liu, Amir Shahroudy, Dong Xu, Gang Wang. Spatio-Temporal LSTM with Trust Gates for 3D Human Action Recognition 816-833. [Crossref]

1317. Shaolei Wang, Wanxiang Che, Yijia Liu, Ting Liu. Enhancing Neural Disfluency Detection with Hand-Crafted Features 336-347. [Crossref]

1318. Lisa Anne Hendricks, Zeynep Akata, Marcus Rohrbach, Jeff Donahue, Bernt Schiele, Trevor Darrell. Generating Visual Explanations 3-19. [Crossref]

1319. Alexander Dörr, Sebastian Otte, Andreas Zell. Investigating Recurrent Neural Networks for Feature-Less Computational Drug Design 140-148. [Crossref]

1320. Da Pan, Meishan Zhang, Guohong Fu. Chinese Sentiment Analysis Exploiting Heterogeneous Segmentations 387-398. [Crossref]

1321. Thorsten Wuest, Daniel Weimer, Christopher Irgens, Klaus-Dieter Thoben. 2016. Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research* 4:1, 23-45. [Crossref]

1322. Ankit Gandhi, Arjun Sharma, Arijit Biswas, Om Deshmukh. GeThR-Net: A Generalized Temporally Hybrid Recurrent Neural Network for Multimodal Information Fusion 883-899. [Crossref]

1323. Xi Peng, Rogerio S. Feris, Xiaoyu Wang, Dimitris N. Metaxas. A Recurrent Encoder-Decoder Network for Sequential Face Alignment 38-56. [Crossref]

1324. Hugo Jair Escalante, Victor Ponce-Lopez, Jun Wan, Michael A. Riegler, Baiyu Chen, Albert Clapes, Sergio Escalera, Isabelle Guyon, Xavier Baro, Pal Halvorsen, Henning Muller, Martha Larson. ChaLearn Joint Contest on Multimedia Challenges Beyond Visual Analysis: An overview 67-73. [Crossref]

1325. SiLiang Li, Bin Xu, Tong Lee Chung. Definition Extraction with LSTM Recurrent Neural Networks 177-189. [Crossref]

1326. Youyi Cai, Wenming Zheng, Tong Zhang, Qiang Li, Zhen Cui, Jiayin Ye. Video Based Emotion Recognition Using CNN and BRNN 679-691. [Crossref]

1327. Yubo Chen, Shulin Liu, Shizhu He, Kang Liu, Jun Zhao. Event Extraction via Bidirectional Long Short-Term Memory Tensor Neural Networks 190-203. [Crossref]

1328. Qi Li, Tianshi Li, Baobao Chang. Learning Word Sense Embeddings from Word Sense Definitions 224-235. [Crossref]

1329. Lei Xie, Lijuan Wang, Shan Yang. Visual Speech Animation 1-30. [Crossref]

1330. Ying Zeng, Honghui Yang, Yansong Feng, Zheng Wang, Dongyan Zhao. A Convolution BiLSTM Neural Network Model for Chinese Event Extraction 275-287. [Crossref]

1331. Nitish Shirish Keskar, Albert S. Berahas. adaQN: An Adaptive Quasi-Newton Algorithm for Training RNNs 1-16. [Crossref]

1332. Zae Myung Kim, Young-Seob Jeong. TIMEX3 and event extraction using recurrent neural networks 450-453. [Crossref]

1333. Albert Haque, Boya Peng, Zelun Luo, Alexandre Alahi, Serena Yeung, Li Fei-Fei. Towards Viewpoint Invariant 3D Human Pose Estimation 160-177. [Crossref]

1334. Alberto Bartoli, Andrea De Lorenzo, Eric Medvet, Fabiano Tarlao. Your Paper has been Accepted, Rejected, or Whatever: Automatic Generation of Scientific Paper Reviews 19-28. [Crossref]

1335. Huiwei Zhou, Huijie Deng, Long Chen, Yunlong Yang, Chen Jia, Degen Huang. 2016. Exploiting syntactic and semantics information for chemical–disease relation extraction. *Database* **2016**, baw048. [Crossref]

1336. Lydia Lazib, Yanyan Zhao, Bing Qin, Ting Liu. Negation Scope Detection with Recurrent Neural Networks Models in Review Texts 494-508. [Crossref]

1337. Bruno Stuner, Clement Chatelain, Thierry Paquet. Cascading BLSTM networks for handwritten word recognition 3416-3421. [Crossref]

1338. Yushi Yao, Zheng Huang. Bi-directional LSTM Recurrent Neural Network for Chinese Word Segmentation 345-353. [Crossref]

1339. Zhen Zuo, Bing Shuai, Wang Gang, Xiao Liu, Xingxing Wang, Bing Wang, Yushi Chen. 2016. Learning Contextual Dependencies with Convolutional Hierarchical Recurrent Neural Networks. *IEEE Transactions on Image Processing* 1-1. [Crossref]

1340. Bálint Pál Tóth, Balázs Szórádi, Géza Németh. Improvements to Prosodic Variation in Long Short-Term Memory Based Intonation Models Using Random Forest 386-394. [Crossref]

1341. Masashi Tsubaki, Masashi Shimbo, Yuji Matsumoto. 2016. Non-linear Similarity Learning for Semantic Compositionality. *Transactions of the Japanese Society for Artificial Intelligence* **31**:2, O-FA2_1-10. [Crossref]

1342. Hao Wang, Dit-Yan Yeung. 2016. Towards Bayesian Deep Learning: A Framework and Some Existing Methods. *IEEE Transactions on Knowledge and Data Engineering* 1-1. [Crossref]

1343. Xiaodan Liang, Xiaohui Shen, Jiashi Feng, Liang Lin, Shuicheng Yan. Semantic Object Parsing with Graph LSTM 125-143. [Crossref]

1344. Giulio Ruffini, David Ibañez, Marta Castellano, Stephen Dunne, Aureli Soria-Frisch. EEG-driven RNN Classification for Prognosis of Neurodegeneration in At-Risk Patients 306-313. [Crossref]

1345. Sulagna Gope, Sudeshna Sarkar, Pabitra Mitra, Subimal Ghosh. Early Prediction of Extreme Rainfall Events: A Deep Learning Approach 154-167. [Crossref]

1346. Natalia Neverova, Christian Wolf, Griffin Lacey, Lex Fridman, Deepak Chandra, Brandon Barbello, Graham Taylor. 2016. Learning Human Identity From Motion Patterns. *IEEE Access* **4**, 1810-1820. [Crossref]

1347. Yang Feng, Yuncheng Li, Jiebo Luo. Learning effective Gait features using LSTM 325-330. [Crossref]

1348. Michimasa Inaba, Kenichi Takahashi. Backchanneling via Twitter Data for Conversational Dialogue Systems 148-155. [Crossref]

1349. Varun K. Nagaraja, Vlad I. Morariu, Larry S. Davis. Modeling Context Between Objects for Referring Expression Understanding 792-807. [Crossref]

1350. Jan Svoboda, Jonathan Masci, Michael M. Bronstein. Palmprint recognition via discriminative index learning 4232-4237. [Crossref]

1351. Naifan Zhuang, Jun Ye, Kien A. Hua. DLSTM approach to video modeling with hashing for large-scale video retrieval 3222-3227. [Crossref]

1352. Liang Wang, Sujian Li, Xinyan Xiao, Yajuan Lyu. Topic Segmentation of Web Documents with Automatic Cue Phrase Identification and BLSTM-CNN 177-188. [Crossref]

1353. Tobias Weyand, Ilya Kostrikov, James Philbin. PlaNet - Photo Geolocation with Convolutional Neural Networks 37-55. [Crossref]

1354. Arthi Venkataraman. Deep learning algorithms based text classifier 220-224. [Crossref]

1355. Bernardino Romera-Paredes, Philip Hilaire Sean Torr. Recurrent Instance Segmentation 312-329. [Crossref]

1356. Qianrong Zhou, Long Ma, Zhenyu Zheng, Yue Wang, Xiaojie Wang. Recurrent Neural Word Segmentation with Tag Inference 734-743. [Crossref]

1357. Michalis Agathocleous, Chris Christodoulou, Vasilis Promponas, Petros Kountouris, Vassilis Vassiliades. Training Bidirectional Recurrent Neural Network Architectures with the Scaled Conjugate Gradient Algorithm 123-131. [Crossref]

1358. Jonas Uhrig, Marius Cordts, Uwe Franke, Thomas Brox. Pixel-Level Encoding and Depth Layering for Instance-Level Semantic Labeling 14-25. [Crossref]

1359. Lasguido NIO, Sakriani SAKTI, Graham NEUBIG, Koichiro YOSHINO, Satoshi NAKAMURA. 2016. Neural Network Approaches to Dialog Response Retrieval and Generation. *IEICE Transactions on Information and Systems* **E99.D**:10, 2508-2517. [Crossref]

1360. Chuang Gan, Chen Sun, Lixin Duan, Boqing Gong. Webly-Supervised Video Recognition by Mutually Voting for Relevant Web Images and Web Video Frames 849-866. [Crossref]

1361. Irman Abdic, Lex Fridman, Daniel E. Brown, William Angell, Bryan Reimer, Erik Marchi, Bjorn Schuller. Detecting road surface wetness from audio: A deep learning approach 3458-3463. [Crossref]

1362. Fred Highland, Corey Hart. 2016. Unsupervised Learning of Patterns Using Multilayer Reverberating Configurations of Polychronous Wavefront Computation. *Procedia Computer Science* **95**, 175-184. [Crossref]

1363. Houssem Maghrebi, Thibault Portigliatti, Emmanuel Prouff. Breaking Cryptographic Implementations Using Deep Learning Techniques 3-26. [Crossref]

1364. Jürgen Schmidhuber. Deep Learning 1-11. [Crossref]

1365. Yili Zhao, Dan Xu. Monet-Style Images Generation Using Recurrent Neural Networks 205-211. [Crossref]

1366. Rania Maalej, Monji Kherallah. Improving MDLSTM for Offline Arabic Handwriting Recognition Using Dropout at Different Positions 431-438. [Crossref]

1367. Zheng Xiao, PiJun Liang. Chinese Sentiment Analysis Using Bidirectional LSTM with Word Embedding 601-610. [Crossref]

1368. Saad Bin Ahmed, Saeeda Naz, Muhammad Imran Razzak, Rubiyah Yusof, Thomas M. Breuel. Balinese Character Recognition Using Bidirectional LSTM Classifier 201-211. [Crossref]

1369. Shengtao Xiao, Jiashi Feng, Junliang Xing, Hanjiang Lai, Shuicheng Yan, Ashraf Kassim. Robust Facial Landmark Detection via Recurrent Attentive-Refinement Networks 57-72. [Crossref]

1370. Armand Joulin, Laurens van der Maaten, Allan Jabri, Nicolas Vasilache. Learning Visual Features from Large Weakly Supervised Data 67-84. [Crossref]

1371. Xinze Chen, Guangliang Cheng, Yinghao Cai, Dayong Wen, Heping Li. Semantic Segmentation with Modified Deep Residual Networks 42-54. [Crossref]

1372. Hyun-Soo Choi, Byunghan Lee, Sungroh Yoon. 2016. Biometric Authentication Using Noisy Electrocardiograms Acquired by Mobile Sensors. *IEEE Access* **4**, 1266-1273. [Crossref]

1373. Kuo-Hao Zeng, Tseng-Hung Chen, Juan Carlos Niebles, Min Sun. Title Generation for User Generated Videos 609-625. [Crossref]

1374. Nan Zhou, Jun Du. Recognition of Social Touch Gestures Using 3D Convolutional Neural Networks 164-173. [Crossref]

1375. Fan Zhu, Jin Xie, Yi Fang. **9911**, 305. [Crossref]

1376. Shyam Sundar Rajagopalan, Louis-Philippe Morency, Tadas Baltrušaitis, Roland Goecke. **9911**, 338. [Crossref]

1377. Tayfun Alpay, Stefan Heinrich, Stefan Wermter. Learning Multiple Timescales in Recurrent Neural Networks 132-139. [Crossref]

1378. Adam P. Trischler, Gabriele M.T. D'Eleuterio. 2016. Synthesis of recurrent neural networks for dynamical system simulation. *Neural Networks* **80**, 67. [Crossref]

1379. Jian Fu, Xipeng Qiu, Xuanjing Huang. Convolutional Deep Neural Networks for Document-Based Question Answering 790-797. [Crossref]

1380. Pengyuan Lyu, Baoguang Shi, Chengquan Zhang, Xiang Bai. Distinguishing text/non-text natural images with Multi-Dimensional Recurrent Neural Networks 3981-3986. [Crossref]

1381. Ke Zhang, Wei-Lun Chao, Fei Sha, Kristen Grauman. Video Summarization with Long Short-Term Memory 766-782. [Crossref]

1382. Milos Manic, Kasun Amarasinghe, Juan J. Rodriguez-Andina, Craig Rieger. 2016. Intelligent Buildings of the Future: Cyberaware, Deep Learning Powered, and Human Interacting. *IEEE Industrial Electronics Magazine* **10**:4, 32-49. [Crossref]

1383. Rania Maalej, Najiba Tagougui, Monji Kherallah. Recognition of Handwritten Arabic Words with Dropout Applied in MDLSTM 746-752. [Crossref]

1384. Yichao Yan, Bingbing Ni, Zhichao Song, Chao Ma, Yan Yan, Xiaokang Yang. **9910**, 701. [Crossref]

1385. Kim Berninger, Jannis Hoppe, Benjamin Milde. Classification of Speaker Intoxication Using a Bidirectional Recurrent Neural Network 435-442. [Crossref]

1386. Raymond Smith, Chunhui Gu, Dar-Shyang Lee, Huiyi Hu, Ranjith Unnikrishnan, Julian Ibarz, Sacha Arnoud, Sophia Lin. End-to-End Interpretation of the French Street Name Signs Dataset 411-426. [Crossref]

1387. Simon Andermatt, Simon Pezold, Philippe Cattin. Multi-dimensional Gated Recurrent Units for the Segmentation of Biomedical 3D-Data 142-151. [Crossref]

1388. Zecheng Xie, Zenghui Sun, Lianwen Jin, Ziyong Feng, Shuye Zhang. Fully convolutional recurrent network for handwritten Chinese text recognition 4011-4016. [Crossref]

1389. Aleksandr Sboev, Tatiana Litvinova, Dmitry Gudovskikh, Roman Rybka, Ivan Moloshnikov. 2016. Machine Learning Models of Text Categorization by Author Gender Using Topic-independent Features. *Procedia Computer Science* **101**, 135-142. [Crossref]

1390. Jun Deng, Nicholas Cummins, Jing Han, Xinzhou Xu, Zhao Ren, Vedhas Pandit, Zixing Zhang, Björn Schuller. The University of Passau Open Emotion Recognition System for the Multimodal Emotion Challenge 652-666. [Crossref]

1391. Jerome R. Bellegarda, Christof Monz. 2016. State of the art in statistical methods for language and speech processing. *Computer Speech & Language* **35**, 163-184. [Crossref]

1392. Junguo Zhu, Muyun Yang, Sheng Li, Tiejun Zhao. Learning Bilingual Sentence Representations for Quality Estimation of Machine Translation 35-42. [Crossref]

1393. Rahul Rama Varior, Mrinal Haloi, Gang Wang. Gated Siamese Convolutional Neural Network Architecture for Human Re-identification 791-808. [Crossref]

1394. Jana Straková, Milan Straka, Jan Hajič. Neural Networks for Featureless Named Entity Recognition in Czech 173-181. [Crossref]

1395. Mari Ostendorf. Continuous-Space Language Processing: Beyond Word Embeddings 3-15. [Crossref]

1396. Florian Eyben. Real-time Incremental Processing 139-161. [Crossref]

1397. Yong Du, Yun Fu, liang wang. 2016. Representation Learning of Temporal Dynamics for Skeleton Based Action Recognition. *IEEE Transactions on Image Processing* 1-1. [Crossref]

1398. Anna Rohrbach, Marcus Rohrbach, Ronghang Hu, Trevor Darrell, Bernt Schiele. Grounding of Textual Phrases in Images by Reconstruction 817-834. [Crossref]

1399. Sheng-Hung Hu, Yikang Li, Baoxin Li. Video2vec: Learning semantic spatio-temporal embeddings for video representation 811-816. [Crossref]

1400. Florian Eyben. Real-Life Robustness 163-183. [Crossref]

1401. Allan Jabri, Armand Joulin, Laurens van der Maaten. Revisiting Visual Question Answering Baselines 727-739. [Crossref]

1402. Yuuki Tachioka, Jun Ishii. 2016. Long short-term memory recurrent-neural-network-based bandwidth extension for automatic speech recognition. *Acoustical Science and Technology* 37:6, 319-321. [Crossref]

1403. Debajyoti Datta, Valentina Brashers, John Owen, Casey White, Laura E. Barnes. A Deep Learning Methodology for Semantic Utterance Classification in Virtual Human Dialogue Systems 451-455. [Crossref]

1404. James Ryan, Adam James Summerville, Michael Mateas, Noah Wardrip-Fruin. Translating Player Dialogue into Meaning Representations Using LSTMs 383-386. [Crossref]

1405. Qingsong Ma, Fandong Meng, Daqi Zheng, Mingxuan Wang, Yvette Graham, Wenbin Jiang, Qun Liu. MaxSD: A Neural Machine Translation Evaluation Metric Optimized by Maximizing Similarity Distance 153-161. [Crossref]

1406. Jun Guo, Hongyang Chao. Building Dual-Domain Representations for Compression Artifacts Reduction 628-644. [Crossref]

1407. Qianrong Zhou, Liyun Wen, Xiaojie Wang, Long Ma, Yue Wang. A Hierarchical LSTM Model for Joint Tasks 324-335. [Crossref]

1408. Zhengkuan Zhang, Weiran Xu, Qianqian Chen. Joint Event Extraction Based on Skip-Window Convolutional Neural Networks 324-334. [Crossref]

1409. Jinyu Li, Li Deng, Reinhold Haeb-Umbach, Yifan Gong. Reverberant speech recognition 203-238. [Crossref]

1410. Artem Chernodub, Dimitri Nowicki. Sampling-Based Gradient Regularization for Capturing Long-Term Dependencies in Recurrent Neural Networks 90-97. [Crossref]

1411. Gunnar A. Sigurdsson, Xinlei Chen, Abhinav Gupta. Learning Visual Storylines with Skipping Recurrent Neural Networks 71-88. [Crossref]

1412. Yanghao Li, Cuiling Lan, Junliang Xing, Wenjun Zeng, Chunfeng Yuan, Jiaying Liu. Online Human Action Detection Using Joint Classification-Regression Recurrent Neural Networks 203-220. [Crossref]

1413. Álvaro Peris, Marc Bolaños, Petia Radeva, Francisco Casacuberta. Video Description Using Bidirectional Recurrent Neural Networks 3-11. [Crossref]

1414. Xiao Lin, Devi Parikh. Leveraging Visual Question Answering for Image-Caption Ranking 261-277. [Crossref]

1415. Ishan Misra, C. Lawrence Zitnick, Martial Hebert. Shuffle and Learn: Unsupervised Learning Using Temporal Order Verification 527-544. [Crossref]

1416. Robert DiPietro, Colin Lea, Anand Malpani, Narges Ahmidi, S. Swaroop Vedula, Gyusung I. Lee, Mija R. Lee, Gregory D. Hager. Recognizing Surgical Activities with Recurrent Neural Networks 551-558. [Crossref]

1417. Jinchao Zhang, Fandong Meng, Mingxuan Wang, Daqi Zheng, Wenbin Jiang, Qun Liu. Is Local Window Essential for Neural Network Based Chinese Word Segmentation? 450-457. [Crossref]

1418. Jiachen Du, Lin Gui, Ruifeng Xu. Extracting Opinion Expression with Neural Attention 151-161. [Crossref]

1419. Taweesak Emsawas, Boonserm Kijsirikul. Thai Printed Character Recognition Using Long Short-Term Memory and Vertical Component Shifting 106-115. [Crossref]

1420. Kai Chen, Qiang Huo. 2016. Training deep bidirectional LSTM acoustic model for LVCSR by a Context-Sensitive-Chunk BPTT approach. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 1-1. [Crossref]

1421. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Identity Mappings in Deep Residual Networks 630-645. [Crossref]

1422. Alp Guler, Nikolaos Kardaris, Siddhartha Chandra, Vassilis Pitsikalis, Christian Werner, Klaus Hauer, Costas Tzafestas, Petros Maragos, Iasonas Kokkinos. Human Joint Angle Estimation and Gesture Recognition for Assistive Robotic Vision 415-431. [Crossref]

1423. Xianbiao Qi, Chun-Guang Li, Guoying Zhao, Xiaopeng Hong, Matti Pietikäinen. 2016. Dynamic texture and scene classification by transferring deep image features. *Neurocomputing* 171, 1230-1241. [Crossref]

1424. Chao Ma, Yun Gu, Wei Liu, Jie Yang, Xiangjian He. Unsupervised Video Hashing by Exploiting Spatio-Temporal Feature 511-518. [Crossref]

1425. Ivan Medennikov, Anna Bulusheva. LSTM-Based Language Models for Spontaneous Speech Recognition 469-475. [Crossref]

1426. Su Jeong Choi, A-Yeong Kim, Seong-Bae Park, Se-Young Park. A Generative Deep Learning for Generating Korean Abbreviations 113-124. [Crossref]

1427. Adela-Diana Almási, Stanisław Woźniak, Valentin Cristea, Yusuf Leblebici, Ton Engbersen. 2016. Review of advances in neural networks: Neural design technology stack. *Neurocomputing* **174**, 31-41. [Crossref]

1428. Kha Gia Quach, Chi Nhan Duong, Khoa Luu, Tien D. Bui. Depth-based 3D hand pose tracking 2746-2751. [Crossref]

1429. Jinwon An, Sungzoon Cho. Hand motion identification of grasp-and-lift task from electroencephalography recordings using recurrent neural networks 427-429. [Crossref]

1430. S.U. Park, J.H. Park, M.A. Al-masni, M.A. Al-antari, Md.Z. Uddin, T.-S. Kim. 2016. A Depth Camera-based Human Activity Recognition via Deep Learning Recurrent Neural Network for Health and Social Care Services. *Procedia Computer Science* **100**, 78-84. [Crossref]

1431. Dongxu Shao, Tianyou Zhang, Kamal Mannar, Yue Han. Time Series Forecasting on Engineering Systems Using Recurrent Neural Networks 459-471. [Crossref]

1432. Peng Wang, Bo Xu, Jiaming Xu, Guanhua Tian, Cheng-Lin Liu, Hongwei Hao. 2016. Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification. *Neurocomputing* **174**, 806-814. [Crossref]

1433. Yanyan Jia, Yansong Feng, Bingfeng Luo, Yuan Ye, Tianyang Liu, Dongyan Zhao. Transition-Based Discourse Parsing with Multilayer Stack Long Short Term Memory 360-373. [Crossref]

1434. Christopher B. Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, Silvio Savarese. 3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction 628-644. [Crossref]

1435. Jiren Jin, Hideki Nakayama. Annotation order matters: Recurrent Image Annotator for arbitrary length image tagging 2452-2457. [Crossref]

1436. Alexander Popov. Deep Learning Architecture for Part-of-Speech Tagging with Word and Suffix Embeddings 68-77. [Crossref]

1437. Narynov Sergazy Sakenovich, Arman Serikuly Zharmagambetov. On One Approach of Solving Sentiment Analysis Task for Kazakh and Russian Languages Using Deep Learning 537-545. [Crossref]

1438. Zhen Li, Yukang Gan, Xiaodan Liang, Yizhou Yu, Hui Cheng, Liang Lin. LSTM-CF: Unifying Context Modeling and Fusion with LSTMs for RGB-D Scene Labeling 541-557. [Crossref]

1439. Marvin Coto-Jiménez, John Goddard-Close, Fabiola Martínez-Licona. Improving Automatic Speech Recognition Containing Additive Noise Using Deep Denoising Autoencoders of LSTM Networks 354-361. [Crossref]

1440. Ronghang Hu, Marcus Rohrbach, Trevor Darrell. Segmentation from Natural Language Expressions 108-124. [Crossref]

1441. Marvin Coto-Jiménez, John Goddard-Close. LSTM Deep Neural Networks Postfiltering for Improving the Quality of Synthetic Voices 280-289. [Crossref]

1442. Maedeh Aghaei, Mariella Dimiccoli, Petia Radeva. With whom do I interact? Detecting social interactions in egocentric photo-streams 2959-2964. [Crossref]

1443. Tianyuan Yu, Liang Bai, Jinlin Guo, Zheng Yang, Yuxiang Xie. A Deep Two-Stream Network for Bidirectional Cross-Media Information Retrieval 328-337. [Crossref]

1444. Mrwan Margem, Ozgur Yilmaz. How Much Computation and Distributedness is Needed in Sequence Learning Tasks? 274-283. [Crossref]

1445. Yuanpu Xie, Zizhao Zhang, Manish Sapkota, Lin Yang. Spatial Clockwork Recurrent Neural Network for Muscle Perimysium Segmentation 185-193. [Crossref]

1446. Loïc Bontemps, Van Loi Cao, James McDermott, Nhien-An Le-Khac. Collective Anomaly Detection Based on Long Short-Term Memory Recurrent Neural Networks 141-152. [Crossref]

1447. Christopher Conly, Alex Dillhoff, Vassilis Athitsos. Leveraging intra-class variations to improve large vocabulary gesture recognition 907-912. [Crossref]

1448. Fei Yan, Krystian Mikolajczyk, Josef Kittler. Generating commentaries for tennis videos 2658-2663. [Crossref]

1449. . References 213-228. [Crossref]

1450. Ozgur Yilmaz. 2015. Symbolic Computation Using Cellular Automata-Based Hyperdimensional Computing. *Neural Computation* **27**:12, 2661-2692. [Abstract] [Full Text] [PDF] [PDF Plus]

1451. Cheng-Tao Chung, Cheng-Yu Tsai, Hsiang-Hung Lu, Chia-Hsiang Liu, Hung-yi Lee, Lin-shan Lee. An iterative deep learning framework for unsupervised discovery of speech features and linguistic units with applications on spoken term detection 245-251. [Crossref]

1452. Junhua Mao, Xu Wei, Yi Yang, Jiang Wang, Zhiheng Huang, Alan L. Yuille. Learning Like a Child: Fast Novel Visual Concept Learning from Sentence Descriptions of Images 2533-2541. [Crossref]

1453. Mateusz Malinowski, Marcus Rohrbach, Mario Fritz. Ask Your Neurons: A Neural-Based Approach to Answering Questions about Images 1-9. [Crossref]

1454. Yangyang Shi, Kaisheng Yao, Hu Chen, Yi-Cheng Pan, Mei-Yuh Hwang. Semi-supervised slot tagging in spoken language understanding using recurrent transductive support vector machines 353-360. [Crossref]

1455. Po-Sen Huang, Minje Kim, Mark Hasegawa-Johnson, Paris Smaragdis. 2015. Joint Optimization of Masks and Deep Recurrent Neural Networks for Monaural Source Separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **23**:12, 2136-2147. [Crossref]

1456. Jinyu Li, Abdelrahman Mohamed, Geoffrey Zweig, Yifan Gong. LSTM time and frequency recurrence for automatic speech recognition 187-191. [Crossref]

1457. Andrew, Hasim Sak, Felix de Chaumont Quitry, Tara Sainath, Kanishka Rao. Acoustic modelling with CD-CTC-SMBR LSTM RNNS 604-609. [Crossref]

1458. Suman Ravuri, Andreas Stoicke. A comparative study of neural network models for lexical intent classification 368-374. [Crossref]

1459. Yifeng Li, Chih-yu Chen, Alice M. Kaye, Wyeth W. Wasserman. 2015. The identification of cis-regulatory elements: A review from a machine learning perspective. *Biosystems* **138**, 6-17. [Crossref]

1460. Yasser Roudi, Graham Taylor. 2015. Learning with hidden variables. *Current Opinion in Neurobiology* **35**, 110-118. [Crossref]

1461. Huan Yang, Baoyuan Wang, Stephen Lin, David Wipf, Minyi Guo, Baining Guo. Unsupervised Extraction of Video Highlights via Robust Recurrent Auto-Encoders 4633-4641. [Crossref]

1462. David Vandyke, Pei-Hao Su, Milica Gasic, Nikola Mrksic, Tsung-Hsien Wen, Steve Young. Multi-domain dialogue success classifiers for policy training 763-770. [Crossref]

1463. Luca Pasa, Alberto Testolin, Alessandro Sperduti. 2015. Neural Networks for Sequential Data: a Pre-training Approach based on Hidden Markov Models. *Neurocomputing* **169**, 323-333. [Crossref]

1464. Sakriani Sakti, Faiz Ilham, Graham Neubig, Tomoki Toda, Ayu Purwarianti, Satoshi Nakamura. Incremental sentence compression using LSTM recurrent networks 252-258. [Crossref]

1465. Xu Jia, Efstratios Gavves, Basura Fernando, Tinne Tuytelaars. Guiding the Long-Short Term Memory Model for Image Caption Generation 2407-2415. [Crossref]

1466. Yongxue Tian, Li Pan. Predicting Short-Term Traffic Flow by Long Short-Term Memory Recurrent Neural Network 153-158. [Crossref]

1467. Lukas Mauch, Bin Yang. A new approach for supervised power disaggregation by using a deep recurrent LSTM network 63-67. [Crossref]

1468. Yoshitaka Ushiku, Masataka Yamaguchi, Yusuke Mukuta, Tatsuya Harada. Common Subspace for Model and Similarity: Phrase Learning for Caption Generation from Images 2668-2676. [Crossref]

1469. Taesup Moon, Heeyoul Choi, Hoshik Lee, Inchul Song. RNNDROP: A novel dropout for RNNS in ASR 65-70. [Crossref]

1470. Jimmy Lei Ba, Kevin Swersky, Sanja Fidler, Ruslan Salakhutdinov. Predicting Deep Zero-Shot Convolutional Neural Networks Using Textual Descriptions 4247-4255. [Crossref]

1471. Anurag Kishore, Sanjay Singh. Natural language image descriptor 110-115. [Crossref]

1472. Yajie Miao, Mohammad Gowayyed, Florian Metze. EESEN: End-to-end speech recognition using deep RNN models and WFST-based decoding 167-174. [Crossref]

1473. Li Yao, Atousa Torabi, Kyunghyun Cho, Nicolas Ballas, Christopher Pal, Hugo Larochelle, Aaron Courville. Describing Videos by Exploiting Temporal Structure 4507-4515. [Crossref]

1474. Katerina Fragkiadaki, Sergey Levine, Panna Felsen, Jitendra Malik. Recurrent Network Models for Human Dynamics 4346-4354. [Crossref]

1475. Bharat Singh, Soham De, Yangmuzi Zhang, Thomas Goldstein, Gavin Taylor. Layer-Specific Adaptive Learning Rates for Deep Networks 364-368. [Crossref]

1476. Chuang Ding, Lei Xie, Jie Yan, Weini Zhang, Yang Liu. Automatic prosody prediction for Chinese speech synthesis using BLSTM-RNN and embedding features 98-102. [Crossref]

1477. Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, Sanja Fidler. Aligning Books and Movies: Towards Story-Like Visual Explanations by Watching Movies and Reading Books 19-27. [Crossref]

1478. Jacob Walker, Abhinav Gupta, Martial Hebert. Dense Optical Flow Prediction from a Static Image 2443-2451. [Crossref]

1479. Lukas Zilka, Filip Jurcicek. Incremental LSTM-based dialog state tracker 757-762. [Crossref]

1480. Subhashini Venugopalan, Marcus Rohrbach, Jeffrey Donahue, Raymond Mooney, Trevor Darrell, Kate Saenko. Sequence to Sequence -- Video to Text 4534-4542. [Crossref]

1481. Vivek Veeriah, Naifan Zhuang, Guo-Jun Qi. Differential Recurrent Neural Networks for Action Recognition 4041-4049. [Crossref]

1482. Chunshui Cao, Xianming Liu, Yi Yang, Yinan Yu, Jiang Wang, Zilei Wang, Yongzhen Huang, Liang Wang, Chang Huang, Wei Xu, Deva Ramanan, Thomas S. Huang. Look and Think Twice: Capturing Top-Down Visual Attention with Feedback Convolutional Neural Networks 2956-2964. [Crossref]

1483. Oliver Nina, Andres Rodriguez. Simplified LSTM unit and search space probability exploration for image description 1-5. [Crossref]

1484. Zhou Yu, Vikram Ramanarayanan, David Suendermann-Oeft, Xinhao Wang, Klaus Zechner, Lei Chen, Jidong Tao, Aliaksei Ivanou, Yao Qian. Using bidirectional lstm recurrent neural networks to learn high-level abstractions of sequential features for automated scoring of non-native spontaneous speech 338-345. [Crossref]

1485. Abdel-rahman Mohamed, Frank Seide, Dong Yu, Jasha Droppo, Andreas Stoicke, Geoffrey Zweig, Gerald Penn. Deep bi-directional recurrent networks over spectral windows 78-83. [Crossref]

1486. Nikolaus Kriegeskorte. 2015. Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. *Annual Review of Vision Science* **1**:1, 417-446. [Crossref]

1487. William H. Alexander, Joshua W. Brown. 2015. Hierarchical Error Representation: A Computational Model of Anterior Cingulate and Dorsolateral Prefrontal Cortex. *Neural Computation* **27**:11, 2354-2410. [Abstract] [Full Text] [PDF] [PDF Plus] [Supplemental Material]

1488. Zhenqi Xu, Shan Li, Weihong Deng. Learning temporal features using LSTM-CNN architecture for face anti-spoofing 141-145. [Crossref]

1489. Cuong Tuan Nguyen, Masaki Nakagawa. An improved segmentation of online English handwritten text using recurrent neural networks 176-180. [Crossref]

1490. Mi Fei, Dit-Yan Yeung. Temporal Models for Predicting Student Dropout in Massive Open Online Courses 256-263. [Crossref]

1491. Takayuki Osogami, Makoto Otsuka. 2015. Seven neurons memorizing sequences of alphabetical images via spike-timing dependent plasticity. *Scientific Reports* **5**:1. . [Crossref]

1492. Kyunghyun Cho, Aaron Courville, Yoshua Bengio. 2015. Describing Multimedia Content Using Attention-Based Encoder-Decoder Networks. *IEEE Transactions on Multimedia* **17**:11, 1875-1886. [Crossref]

1493. Fabien Ringeval, Florian Eyben, Eleni Kroupi, Anil Yuce, Jean-Philippe Thiran, Touradj Ebrahimi, Denis Lalanne, Björn Schuller. 2015. Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data. *Pattern Recognition Letters* **66**, 22-30. [Crossref]

1494. Sanjay Bilakhia, Stavros Petridis, Anton Nijholt, Maja Pantic. 2015. The MAHNOB Mimicry Database: A database of naturalistic human interactions. *Pattern Recognition Letters* **66**, 52-61. [Crossref]

1495. Tobias Brosch, Heiko Neumann, Pieter R. Roelfsema. 2015. Reinforcement Learning of Linking and Tracing Contours in Recurrent Neural Networks. *PLOS Computational Biology* **11**:10, e1004489. [Crossref]

1496. Keiichi Osako, Rita Singh, Bhiksha Raj. Complex recurrent neural networks for denoising speech signals 1-5. [Crossref]

1497. Alan Ferrari, Daniele Puccinelli, Silvia Giordano. Gesture-based soft authentication 771-777. [Crossref]

1498. Raunaq Vohra, Kratarth Goel, J. K. Sahoo. Modeling temporal dependencies in data using a DBN-LSTM 1-4. [Crossref]

1499. Kui Wu, Xuancong Wang, Nina Zhou, AiTi Aw, Haizhou Li. Joint Chinese word segmentation and punctuation prediction using deep recurrent neural network for social media data 41-44. [Crossref]

1500. Wonmin Byeon, Marcus Liwicki, Thomas M. Breuel. 2015. Scene analysis by mid-level attribute learning using 2D LSTM networks and an application to web-image tagging. *Pattern Recognition Letters* **63**, 23-29. [Crossref]

1501. Kai Chen, Yi Zhou, Fangyan Dai. A LSTM-based method for stock returns prediction: A case study of China stock market 2823-2824. [Crossref]

1502. Chandra Khatri, Suman Voleti, Sathish Veeraraghavan, Nish Parikh, Atiq Islam, Shifa Mahmood, Neeraj Garg, Vivek Singh. Algorithmic content generation for products 2945-2947. [Crossref]

1503. Marcus Edel, Enrico Koppe. An advanced method for pedestrian dead reckoning using BLSTM-RNNs 1-6. [Crossref]

1504. Urminder Singh, Sucheta Chauhan, A. Krishnamachari, Lovekesh Vig. Ensemble of deep long short term memory networks for labelling origin of replication sequences 1-7. [Crossref]

1505. Sucheta Chauhan, Lovekesh Vig. Anomaly detection in ECG time signals via deep long short-term memory networks 1-7. [Crossref]

1506. Kohei Nakajima, Helmut Hauser, Tao Li, Rolf Pfeifer. 2015. Information processing via physical soft body. *Scientific Reports* **5**:1. . [Crossref]

1507. Francesco Donnarumma, Roberto Prevete, Fabian Chersi, Giovanni Pezzulo. 2015. A Programmer–Interpreter Neural Network Architecture for Prefrontal Cognitive Control. *International Journal of Neural Systems* **25**:06, 1550017. [Crossref]

1508. Leimin Tian, Johanna D. Moore, Catherine Lai. Recognizing emotions in dialogues with acoustic and lexical features 737-742. [Crossref]

1509. Adriana Horelu, Catalin Leordeanu, Elena Apostol, Dan Huru, Mariana Mocanu, Valentin Cristea. Forecasting Techniques for Time Series from Sensor Data 261-264. [Crossref]

1510. Leimin Tian, Johanna D. Moore, Catherine Lai. Emotion recognition in spontaneous and acted dialogues 698-704. [Crossref]

1511. Gereon Buscher, Martin Meier, Guillaume Walck, Robert Haschke, Helge J. Ritter. Augmenting curved robot surfaces with soft tactile skin 1514-1519. [Crossref]

1512. Jianri Li, Jong-Hyeok Lee. 2015. Korean Transition-based Dependency Parsing with Recurrent Neural Network. *KIISE Transactions on Computing Practices* **21**:8, 567-571. [Crossref]

1513. Bolan Su, Xi Zhang, Shijian Lu, Chew Lim Tan. Segmented handwritten text recognition with recurrent neural network classifiers 386-390. [Crossref]

1514. Adnan Ul-Hasan, Muhammad Zeshan Afzal, Faisal Shafait, Marcus Liwicki, Thomas M. Breuel. A sequence learning approach for multiple script identification 1046-1050. [Crossref]

1515. Qi Liu, Lijuan Wang, Qiang Huo. A study on effects of implicit and explicit language model information for DBLSTM-CTC based handwriting recognition 461-465. [Crossref]

1516. Fotini Simistira, Adnan Ul-Hassan, Vassilis Papavassiliou, Basilis Gatos, Vassilis Katsouros, Marcus Liwicki. Recognition of historical Greek polytonic scripts using LSTM networks 766-770. [Crossref]

1517. Bastien Moysset, Christopher Kermorvant, Christian Wolf, Jerome Louradour. Paragraph text segmentation into lines with Recurrent Neural Networks 456-460. [Crossref]

1518. Utpal Garain, Luc Mioulet, B. B. Chaudhuri, Clement Chatelain, Thierry Paquet. Unconstrained Bengali handwriting recognition with recurrent models 1056-1060. [Crossref]

1519. Mahdi Hamdani, Patrick Doetsch, Hermann Ney. Bagging by design for continuous Handwriting Recognition using multi-objective particle swarm optimization 256-260. [Crossref]

1520. Federico Raue, Wonmin Byeon, Thomas M. Breuel, Marcus Liwicki. Parallel sequence classification using recurrent neural networks and alignment 581-585. [Crossref]

1521. Kai Chen, Zhi-Jie Yan, Qiang Huo. A context-sensitive-chunk BPTT approach to training deep LSTM/BLSTM recurrent neural networks for offline handwriting recognition 411-415. [Crossref]

1522. Bernhard Lehner, Gerhard Widmer, Sebastian Bock. A low-latency, real-time-capable singing voice detection method with LSTM recurrent neural networks 21-25. [Crossref]

1523. Mohammad Reza Yousefi, Mohammad Reza Soheili, Thomas M. Breuel, Ehsanollah Kabir, Didier Stricker. Binarization-free OCR for historical documents using LSTM networks 1121-1125. [Crossref]

1524. Adnan Ul-Hasan, Faisal Shafaity, Marcus Liwicki. Curriculum learning for printed text line recognition of ligature-based scripts 1001-1005. [Crossref]

1525. Angelica I. Aviles, Samar M. Alsaleh, Pilar Sobrevilla, Alicia Casals. Force-feedback sensory substitution using supervised recurrent learning for robotic-assisted surgery 1-4. [Crossref]

1526. Theodore Bluche, Christopher Kermorvant, Jerome Louradour. Where to apply dropout in recurrent neural networks for handwriting recognition? 681-685. [Crossref]

1527. Luc Mioulet, Utpal Garain, Clement Chatelain, Philippine Barlas, Thierry Paquet. Language identification from handwritten documents 676-680. [Crossref]

1528. Yang Gao, Guyue Mi, Ying Tan. Variable length concentration based feature construction method for spam detection 1-7. [Crossref]

1529. Giacomo Ferroni, Roberto Bonfigli, Emanuele Principi, Stefano Squartini, Francesco Piazza. A Deep Neural Network approach for Voice Activity Detection in multi-room domestic scenarios 1-8. [Crossref]

1530. Juyang Weng. Brains as naturally emerging turing machines 1-8. [Crossref]

1531. Erik Marchi, Fabio Vesperini, Felix Weninger, Florian Eyben, Stefano Squartini, Bjorn Schuller. Non-linear prediction with LSTM recurrent neural networks for acoustic novelty detection 1-7. [Crossref]

1532. Davide Zambrano, Pieter R. Roelfsema, Sander M. Bohte. Continuous-time on-policy neural Reinforcement Learning of working memory tasks 1-8. [Crossref]

1533. Y. Ollivier. 2015. Riemannian metrics for neural networks II: recurrent networks and learning symbolic data sequences. *Information and Inference* **4**:2, 154-193. [Crossref]

1534. Xinlei Chen, C. Lawrence Zitnick. Mind's eye: A recurrent visual representation for image caption generation 2422-2431. [Crossref]

1535. Peter Foster, Simon Dixon, Anssi Klapuri. 2015. Identifying Cover Songs Using Information-Theoretic Measures of Similarity. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **23**:6, 993-1005. [Crossref]

1536. Kun Han, Yuxuan Wang, DeLiang Wang, William S. Woods, Ivo Merks, Tao Zhang. 2015. Learning Spectral Mapping for Speech Dereverberation and Denoising. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **23**:6, 982-992. [Crossref]

1537. Joe Yue-Hei Ng, Matthew Hausknecht, Sudheendra Vijayanarasimhan, Oriol Vinyals, Rajat Monga, George Toderici. Beyond short snippets: Deep networks for video classification 4694-4702. [Crossref]

1538. Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan. Show and tell: A neural image caption generator 3156-3164. [Crossref]

1539. Gheith A. Abandah, Alex Graves, Balkees Al-Shagoor, Alaa Arabiyat, Fuad Jamour, Majid Al-Taee. 2015. Automatic diacritization of Arabic text using recurrent neural networks. *International Journal on Document Analysis and Recognition (IJDAR)* **18**:2, 183-197. [Crossref]

1540. Wonmin Byeon, Thomas M. Breuel, Federico Raue, Marcus Liwicki. Scene labeling with LSTM recurrent neural networks 3547-3555. [Crossref]

1541. Yann LeCun, Yoshua Bengio, Geoffrey Hinton. 2015. Deep learning. *Nature* **521**:7553, 436-444. [Crossref]

1542. Norbert Michael Mayer. 2015. Input-Anticipating Critical Reservoirs Show Power Law Forgetting of Unexpected Input Events. *Neural Computation* **27**:5, 1102-1119. [Abstract] [Full Text] [PDF] [PDF Plus]

1543. Xiaolei Ma, Zhimin Tao, Yinhai Wang, Haiyang Yu, Yunpeng Wang. 2015. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies* **54**, 187-197. [Crossref]

1544. Yong Shean Chong, Yong Haur Tay. Modeling video-based anomaly detection using deep architectures: Challenges and possibilities 1-8. [Crossref]

1545. Bo Fan, Lijuan Wang, Frank K. Soong, Lei Xie. Photo-real talking head with deep bidirectional LSTM 4884-4888. [Crossref]

1546. Kfir Bar, Nachum Dershowitz, Lior Wolf, Yackov Lubarsky, Yaacov Choueka. Processing Judeo-Arabic Texts 138-144. [Crossref]

1547. Lifa Sun, Shiyin Kang, Kun Li, Helen Meng. Voice conversion using deep Bidirectional Long Short-Term Memory based Recurrent Neural Networks 4869-4873. [Crossref]

1548. Ouais Alsharif, Tom Ouyang, Francoise Beaufays, Shumin Zhai, Thomas Breuel, Johan Schalkwyk. Long short term memory neural network for keyboard gesture decoding 2076-2080. [Crossref]

1549. Ebru Arisoy, Abhinav Sethy, Bhuvana Ramabhadran, Stanley Chen. Bidirectional recurrent neural network language models for automatic speech recognition 5421-5425. [Crossref]

1550. Peng Liu, Quanjie Yu, Zhiyong Wu, Shiyin Kang, Helen Meng, Lianhong Cai. A deep recurrent approach for acoustic-to-articulatory inversion 4450-4454. [Crossref]

1551. Xiangang Li, Xihong Wu. Improving long short-term memory networks using maxout units for large vocabulary speech recognition 4600-4604. [Crossref]

1552. Atsunori Ogawa, Takaaki Hori. ASR error detection and recognition rate estimation using deep bidirectional recurrent neural networks 4370-4374. [Crossref]

1553. Kanishka Rao, Fuchun Peng, Hasim Sak, Francoise Beaufays. Grapheme-to-phoneme conversion using Long Short-Term Memory recurrent neural networks 4225-4229. [Crossref]

1554. Heiga Zen, Hasim Sak. Unidirectional long short-term memory recurrent neural network with recurrent output layer for low-latency speech synthesis 4470-4474. [Crossref]

1555. Razvan Pascanu, Jack W. Stokes, Hermineh Sanossian, Mady Marinescu, Anil Thomas. Malware classification with recurrent networks 1916-1920. [Crossref]

1556. Simon Leglaive, Romain Hennequin, Roland Badeau. Singing voice detection with deep recurrent neural networks 121-125. [Crossref]

1557. Hakan Erdogan, John R. Hershey, Shinji Watanabe, Jonathan Le Roux. Phase-sensitive and recognition-boosted speech separation using deep recurrent neural networks 708-712. [Crossref]

1558. Andrew Senior, Hasim Sak, Izhak Shafran. Context dependent phone models for LSTM RNN acoustic modelling 4585-4589. [Crossref]

1559. Moritz Helmstaedter. 2015. The Mutual Inspirations of Machine Learning and Neuroscience. *Neuron* **86**:1, 25-28. [Crossref]

1560. Daniel Renshaw, Keith B. Hall. Long short-term memory language models with additive morphological features for automatic speech recognition 5246-5250. [Crossref]

1561. Peilu Wang, Yao Qian, Frank K. Soong, Lei He, Hai Zhao. Word embedding for recurrent neural network based TTS synthesis 4879-4883. [Crossref]

1562. Paul Voigtlaender, Patrick Doetsch, Simon Wiesler, Ralf Schluter, Hermann Ney. Sequence-discriminative training of recurrent neural networks 2100-2104. [Crossref]

1563. Xiangang Li, Xihong Wu. Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition 4520-4524. [Crossref]

1564. Xiaolei Ma, Haiyang Yu, Yunpeng Wang, Yinhai Wang. 2015. Large-Scale Transportation Network Congestion Evolution Prediction Using Deep Learning Theory. *PLOS ONE* **10**:3, e0119044. [Crossref]

1565. Jaldert O. Rombouts, Sander M. Bohte, Pieter R. Roelfsema. 2015. How Attention Can Create Synaptic Tags for the Learning of Working Memories in Sequential Tasks. *PLOS Computational Biology* **11**:3, e1004060. [Crossref]

1566. Oliver Faust, Wenwei Yu, U. Rajendra Acharya. 2015. The role of real-time in biomedical science: A meta-analysis on computational complexity, delay and speedup. *Computers in Biology and Medicine* **58**, 73-84. [Crossref]

1567. Martin Sundermeyer, Hermann Ney, Ralf Schluter. 2015. From Feedforward to Recurrent LSTM Neural Networks for Language Modeling. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **23**:3, 517-529. [Crossref]

1568. Wim De Mulder, Steven Bethard, Marie-Francine Moens. 2015. A survey on the application of recurrent neural networks to statistical language modeling. *Computer Speech & Language* **30**:1, 61-98. [Crossref]

1569. Christopher H Chatham, David Badre. 2015. Multiple gates on working memory. *Current Opinion in Behavioral Sciences* **1**, 23-31. [Crossref]

1570. Michael J. Frank, David Badre. 2015. How cognitive theory guides neuroscience. *Cognition* **135**, 14-20. [Crossref]

1571. Felix Weninger, Martin Wöllmer, Björn Schuller. Emotion Recognition in Naturalistic Speech and Language-A Survey 237-267. [Crossref]

1572. Christos D. Katsis, George Rigas, Yorgos Goletsis, Dimitrios I. Fotiadis. Emotion Recognition in Car Industry 515-544. [Crossref]

1573. Qi Lyu, Zhiyong Wu, Jun Zhu. Polyphonic Music Modelling with LSTM-RTRBM 991-994. [Crossref]

1574. Chen Sun, Sanketh Shetty, Rahul Sukthankar, Ram Nevatia. Temporal Localization of Fine-Grained Actions in Videos by Domain Transfer from Web Images 371-380. [Crossref]

1575. David Balderas, Arturo Molina, Pedro Ponce. 2015. Alternative Classification Techniques for Brain-Computer Interfaces for Smart Sensor Manufacturing Environments. *IFAC-PapersOnLine* **48**:3, 680-685. [Crossref]

1576. Shizhe Chen, Qin Jin. Multi-modal Dimensional Emotion Recognition using Recurrent Neural Networks 49-56. [Crossref]

1577. Anna Rohrbach, Marcus Rohrbach, Bernt Schiele. The Long-Short Story of Movie Description 209-221. [Crossref]

1578. Łukasz Brocki, Krzysztof Marasek. 2015. Deep Belief Neural Networks and Bidirectional Long-Short Term Memory Hybrid for Speech Recognition. *Archives of Acoustics* **40**:2. . [Crossref]

1579. Daniel Soutner, Luděk Müller. On Continuous Space Word Representations as Input of LSTM Language Model 267-274. [Crossref]

1580. A.G. Sboev, R. Rybka, I. Moloshnikov, D. Gudovskih. 2015. Syntactic Analysis of the Sentences of the Russian Language Based on Neural Networks. *Procedia Computer Science* **66**, 277-286. [Crossref]

1581. Lukáš Žilka, Filip Jurčíček. LecTrack: Incremental Dialog State Tracking with Long Short-Term Memory Networks 174-182. [Crossref]

1582. Jack Kelly, William Knottenbelt. Neural NILM 55-64. [Crossref]

1583. Alan Ferrari, Daniele Puccinelli, Silvia Giordano. Poster 215-217. [Crossref]

1584. Jürgen Schmidhuber. 2015. Deep learning in neural networks: An overview. *Neural Networks* **61**, 85-117. [Crossref]

1585. Ali A. Minai. Computational Models of Cognitive and Motor Control 665-682. [Crossref]

1586. P. Szymczyk, M. Szymczyk. 2015. Classification of geological structure using ground penetrating radar and Laplace transform artificial neural networks. *Neurocomputing* **148**, 354-362. [Crossref]

1587. Amjad Almahairi, Kyle Kastner, Kyunghyun Cho, Aaron Courville. Learning Distributed Representations from Reviews for Collaborative Filtering 147-154. [Crossref]

1588. Baolin Peng, Kaisheng Yao, Li Jing, Kam-Fai Wong. Recurrent Neural Networks with External Memory for Spoken Language Understanding 25-35. [Crossref]

1589. Alessandro Sordoni, Yoshua Bengio, Hossein Vahabi, Christina Lioma, Jakob Grue Simonsen, Jian-Yun Nie. A Hierarchical Recurrent Encoder-Decoder for Generative Context-Aware Query Suggestion 553-562. [Crossref]

1590. Peter Tino, Lubica Benuskova, Alessandro Sperduti. Artificial Neural Network Models 455-471. [Crossref]

1591. Aleksandar Milchevski, Alessandro Rozza, Dimitar Taskovski. Multimodal Affective Analysis combining Regularized Linear Regression and Boosted Regression Trees 33-39. [Crossref]

1592. Xiaodan Zhang, Xinhang Song, Xiong Lv, Shuqiang Jiang, Qixiang Ye, Jianbin Jiao. Rich Image Description Based on Regions 1315-1318. [Crossref]

1593. Mohammad Soleymani, Sadjad Asghari Esfeden, Yun Fu, Maja Pantic. 2015. Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing* 1-1. [Crossref]

1594. Guang Li, Shubo Ma, Yahong Han. Summarization-based Video Caption via Deep Neural Networks 1191-1194. [Crossref]

1595. Anupama Ray, Sai Rajeswar, Santanu Chaudhury. Text recognition using deep BLSTM networks 1-6. [Crossref]

1596. Alan Ferrari, Daniel Puccinelli, Silvia Giordano. Poster 13-15. [Crossref]

1597. Janderson B. Nascimento, Marco Cristo. The Impact of Structured Event Embeddings on Scalable Stock Forecasting Models 121-124. [Crossref]

1598. Zhenzhou Wu, Bao Hong Tan, Rubing Duan, Yong Liu, Rick Siow Mong Goh. Neural Modeling of Buying Behaviour for E-Commerce from Clicking Patterns 1-4. [Crossref]

1599. Frédéric Dandurand, Thomas R. Shultz. 2014. A comprehensive model of development on the balance-scale task. *Cognitive Systems Research* **31-32**, 1-25. [Crossref]

1600. Felix Weninger, John R. Hershey, Jonathan Le Roux, Bjorn Schuller. Discriminatively trained recurrent neural networks for single-channel speech separation 577-581. [Crossref]

1601. Oliver Obst. 2014. Distributed Fault Detection in Sensor Networks using a Recurrent Neural Network. *Neural Processing Letters* **40**:3, 261-273. [Crossref]

1602. Kun Han, DeLiang Wang. 2014. Neural Network Based Pitch Tracking in Very Noisy Speech. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **22**:12, 2158-2168. [Crossref]

1603. Kaisheng Yao, Baolin Peng, Yu Zhang, Dong Yu, Geoffrey Zweig, Yangyang Shi. Spoken language understanding using long short-term memory neural networks 189-194. [Crossref]

1604. Alexandros Gkiokas, Alexandra I. Cristea. Unsupervised neural controller for Reinforcement Learning action-selection: Learning to represent knowledge 99-104. [Crossref]

1605. Adrian Horzyk. 2014. How does generalization and creativity come into being in neural associative systems and how does it form human-like knowledge?. *Neurocomputing* **144**, 238-257. [Crossref]

1606. Alain Droniou, Serena Ivaldi, Olivier Sigaud. Learning a repertoire of actions with deep neural networks 229-234. [Crossref]

1607. Subhrajit Roy, Amitava Banerjee, Arindam Basu. 2014. Liquid State Machine With Dendritically Enhanced Readout for Low-Power, Neuromorphic VLSI Implementations. *IEEE Transactions on Biomedical Circuits and Systems* **8**:5, 681-695. [Crossref]

1608. Bun Theang Ong, Komei Sugiura, Koji Zettsu. Dynamic pre-training of Deep Recurrent Neural Networks for predicting environmental monitoring data 760-765. [Crossref]

1609. Felipe Mendonca Gouveia, Byron Leite Dantas Bezerra, Cleber Zanchettin, Joao Raul Jardim Meneses. Handwriting recognition system for mobile accessibility to the visually impaired people 3918-3923. [Crossref]

1610. Stefan Glüge, Ronald Böck, Günther Palm, Andreas Wendemuth. 2014. Learning long-term dependencies in segmented-memory recurrent neural networks with backpropagation of error. *Neurocomputing* **141**, 54-64. [Crossref]

1611. S. Perrard, M. Labousse, E. Fort, Y. Couder. 2014. Chaos Driven by Interfering Memory. *Physical Review Letters* **113**:10. . [Crossref]

1612. Gheith A. Abandah, Fuad T. Jamour, Esam A. Qaralleh. 2014. Recognizing handwritten Arabic words using grapheme segmentation and recurrent neural networks. *International Journal on Document Analysis and Recognition (IJDAR)* **17**:3, 275-291. [Crossref]

1613. David Daniel Cox, Thomas Dean. 2014. Neural Networks and Neuroscience-Inspired Computer Vision. *Current Biology* **24**:18, R921-R929. [Crossref]

1614. Truyen Van Phan, Masaki Nakagawa. Text/Non-text Classification in Online Handwritten Documents with Recurrent Neural Networks 23-28. [Crossref]

1615. Ryota Ogata, Minoru Mori, Volkmar Frinken, Seiichi Uchida. Constrained AdaBoost for Totally-Ordered Global Features 393-398. [Crossref]

1616. Vu Pham, Theodore Bluche, Christopher Kermorvant, Jerome Louradour. Dropout Improves Recurrent Neural Networks for Handwriting Recognition 285-290. [Crossref]

1617. Bastien Moysset, Ronaldo Messina, Christopher Kermorvant. A Comparison of Recognition Strategies for Printed/Handwritten Composite Documents 158-163. [Crossref]

1618. Patrick Doetsch, Michal Kozielski, Hermann Ney. Fast and Robust Training of Recurrent Neural Networks for Offline Handwriting Recognition 279-284. [Crossref]

1619. Bastien Moysset, Theodore Bluche, Maxime Knibbe, Mohamed Faouzi Benzeghiba, Ronaldo Messina, Jerome Louradour, Christopher Kermorvant. The A2iA Multi-lingual Text Recognition System at the Second Maurdor Evaluation 297-302. [Crossref]

1620. Volkmar Frinken, Yutaro Iwakiri, Ryosuke Ishida, Kensho Fujisaki, Seiichi Uchida. Improving Point of View Scene Recognition by Considering Textual Data 2966-2971. [Crossref]

1621. Wonmin Byeon, Marcus Liwicki, Thomas M. Breuel. Texture Classification Using 2D LSTM Networks 1144-1149. [Crossref]

1622. Geoffrey Hinton. 2014. Where Do Features Come From?. *Cognitive Science* **38**:6, 1078-1101. [Crossref]

1623. Zixing Zhang, Joel Pinto, Christian Plahl, Björn Schuller, Daniel Willett. 2014. Channel mapping using bidirectional long short-term memory for dereverberation in hands-free voice controlled devices. *IEEE Transactions on Consumer Electronics* **60**:3, 525-533. [Crossref]

1624. Matthew M. Botvinick, Jonathan D. Cohen. 2014. The Computational and Neural Basis of Cognitive Control: Charted Territory and New Frontiers. *Cognitive Science* **38**:6, 1249-1285. [Crossref]

1625. Xi Zhang, Chew Lim Tan. Unconstrained Handwritten Word Recognition Based on Trigrams Using BLSTM 2914-2919. [Crossref]

1626. Erik Marchi, Giacomo Ferroni, Florian Eyben, Stefano Squartini, Bjorn Schuller. Audio onset detection: A wavelet packet based approach with recurrent neural networks 3585-3591. [Crossref]

1627. Mohammad Soleymani, Sadjad Asghari-Esfeden, Maja Pantic, Yun Fu. Continuous emotion detection using EEG signals and facial expressions 1-6. [Crossref]

1628. Jia Jun Zhang, Yu Ting Shi. Face recognition systems based on independent component analysis and support vector machine 296-300. [Crossref]

1629. Martin Längkvist, Lars Karlsson, Amy Loutfi. 2014. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters* **42**, 11-24. [Crossref]

1630. Jurgen T. Geiger, Felix Weninger, Jort F. Gemmeke, Martin Wollmer, Bjorn Schuller, Gerhard Rigoll. 2014. Memory-Enhanced Neural Networks and NMF for Robust ASR. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **22**:6, 1037-1046. [Crossref]

1631. Raymond Brueckner, Bjorn Schulter. Social signal classification using deep blstm recurrent neural networks 4823-4827. [Crossref]

1632. Martin Wöllmer, Björn Schuller. 2014. Probabilistic speech feature extraction with context-sensitive Bottleneck neural networks. *Neurocomputing* **132**, 113-120. [Crossref]

1633. Georg Jager, Sebastian Zug, Tino Brade, Andre Dietrich, Christoph Steup, Christian Moewes, Ana-Maria Cretu. Assessing neural networks for sensor fault detection 70-75. [Crossref]

1634. Erik Marchi, Giacomo Ferroni, Florian Eyben, Leonardo Gabrielli, Stefano Squartini, Bjorn Schuller. Multi-resolution linear prediction based features for audio onset detection with bidirectional LSTM neural networks 2164-2168. [Crossref]

1635. Felix Weninger, Shinji Watanabe, Yuuki Tachioka, Bjorn Schuller. Deep recurrent de-noising auto-encoder and blind de-reverberation for reverberated speech recognition 4623-4627. [Crossref]

1636. David Sussillo. 2014. Neural circuits as computational dynamical systems. *Current Opinion in Neurobiology* **25**, 156-163. [Crossref]

1637. Theodore Bluche, Jerome Louradour, Maxime Knibbe, Bastien Moysset, Mohamed Faouzi Benzeghiba, Christopher Kermorvant. The A2iA Arabic Handwritten Text Recognition System at the Open HaRT2013 Evaluation 161-165. [Crossref]

1638. Fabian Triefenbach, Kris Demuynck, Jean-Pierre Martens. 2014. Large Vocabulary Continuous Speech Recognition With Reservoir-Based Acoustic Models. *IEEE Signal Processing Letters* **21**:3, 311-315. [Crossref]

1639. S. F. Feng, M. Schwemmer, S. J. Gershman, J. D. Cohen. 2014. Multitasking versus multiplexing: Toward a normative account of limitations in the simultaneous execution of control-demanding behaviors. *Cognitive, Affective, & Behavioral Neuroscience* **14**:1, 129-146. [Crossref]

1640. Francois Rivest, John F. Kalaska, Yoshua Bengio. 2014. Conditioning and time representation in long short-term memory networks. *Biological Cybernetics* **108**:1, 23-48. [Crossref]

1641. Christopher H. Chatham, Michael J. Frank, David Badre. 2014. Corticostriatal Output Gating during Selection from Working Memory. *Neuron* **81**:4, 930-942. [Crossref]

1642. Marcus Liwicki. Online Handwriting Analysis and Recognition 297-311. [Crossref]

1643. Barbara Hammer. Neural Networks 817-855. [Crossref]

1644. Markus Kächele, Martin Schels, Friedhelm Schwenker. Inferring Depression and Affect from Application Dependent Meta Knowledge 41-48. [Crossref]

1645. C. Otte, S. Otte, L. Wittig, G. Hüttmann, C. Kugler, D. Drömann, A. Zell, A. Schlaefer. 2014. Investigating Recurrent Neural Networks for OCT A-scan Based Tissue Analysis. *Methods of Information in Medicine* **53**:4, 245-249. [Crossref]

1646. Michael Fairbank, Shuhui Li, Xingang Fu, Eduardo Alonso, Donald Wunsch. 2014. An adaptive recurrent neural-network controller using a stabilization matrix and predictive inputs to solve a tracking problem under disturbances. *Neural Networks* **49**, 74-86. [Crossref]

1647. Derek D. Monner, James A. Reggia. 2013. Recurrent Neural Collective Classification. *IEEE Transactions on Neural Networks and Learning Systems* **24**:12, 1932-1943. [Crossref]

1648. Natalia Neverova, Christian Wolf, Giulio Paci, Giacomo Sommavilla, Graham W. Taylor, Florian Nebout. A Multi-scale Approach to Gesture Detection and Recognition 484-491. [Crossref]

1649. Alex Graves, Navdeep Jaitly, Abdel-rahman Mohamed. Hybrid speech recognition with Deep Bidirectional LSTM 273-278. [Crossref]

1650. A. Izady, K. Davary, A. Alizadeh, A. Moghaddam Nia, A. N. Ziaei, S. M. Hasheminia. 2013. Application of NN-ARX Model to Predict Groundwater Levels in the Neishaboor Plain, Iran. *Water Resources Management* **27**:14, 4773-4794. [Crossref]

1651. Fabian Triefenbach, Azarakhsh Jalalvand, Kris Demuynck, Jean-Pierre Martens. 2013. Acoustic Modeling With Hierarchical Reservoirs. *IEEE Transactions on Audio, Speech, and Language Processing* **21**:11, 2439-2450. [Crossref]

1652. Manh Cong Phan, Martin T. Hagan. 2013. Error Surface of Recurrent Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems* **24**:11, 1709-1721. [Crossref]

1653. Erkin Cilden, Faruk Polat. Generating Memoryless Policies Faster Using Automatic Temporal Abstractions for Reinforcement Learning with Hidden State 719-726. [Crossref]

1654. Anna Esposito. The Situated Multimodal Facets of Human Communication 173-202. [Crossref]

1655. . Machine-Based Modelling 235-280. [Crossref]

1656. Sanjay Bilakhia, Stavros Petridis, Maja Pantic. Audiovisual Detection of Behavioural Mimicry 123-128. [Crossref]

1657. S. Otte, C. Otte, A. Schlaefer, L. Wittig, G. Huttmann, D. Dromann, A. Zeli. OCT A-Scan based lung tumor tissue classification with Bidirectional Long Short Term Memory networks 1-6. [Crossref]

1658. Alexander Grushin, Derek D. Monner, James A. Reggia, Ajay Mishra. Robust human action recognition via long short-term memory 1-8. [Crossref]

1659. Tom Schaul. A video game description language for model-based or interactive learning 1-8. [Crossref]

1660. Olivier Morillot, Laurence Likforman-Sulem, Emmanuele Grosicki. Comparative Study of HMM and BLSTM Segmentation-Free Approaches for the Recognition of Handwritten Text-Lines 783-787. [Crossref]

1661. Thomas M. Breuel, Adnan Ul-Hasan, Mayce Ali Al-Azawi, Faisal Shafait. High-Performance OCR for Printed English and Fraktur Using LSTM Networks 683-687. [Crossref]

1662. Branko Šter. 2013. Selective Recurrent Neural Network. *Neural Processing Letters* **38**:1, 1-15. [Crossref]

1663. Thomas Rückstieß, Christian Osendorfer, Patrick van der Smagt. 2013. Minimizing data consumption with sequential online feature selection. *International Journal of Machine Learning and Cybernetics* **4**:3, 235-243. [Crossref]

1664. Juyang Weng, M. D. Luciw, Qi Zhang. 2013. Brain-Like Emergent Temporal Processing: Emergent Open States. *IEEE Transactions on Autonomous Mental Development* **5**:2, 89-116. [Crossref]

1665. Yoshua Bengio, Nicolas Boulanger-Lewandowski, Razvan Pascanu. Advances in optimizing recurrent networks 8624-8628. [Crossref]

1666. Alex Graves, Abdel-rahman Mohamed, Geoffrey Hinton. Speech recognition with deep recurrent neural networks 6645-6649. [Crossref]

1667. Christian Plahl, Michael Kozielski, Ralf Schluter, Hermann Ney. Feature combination and stacking of recurrent and non-recurrent neural networks for LVCSR 6714-6718. [Crossref]

1668. Martin Wollmer, Bjorn Schuller, Gerhard Rigoll. Probabilistic asr feature extraction applying context-sensitive connectionist temporal classification networks 7125-7129. [Crossref]

1669. Florian Eyben, Felix Weninger, Stefano Squartini, Bjorn Schuller. Real-life voice activity detection with LSTM Recurrent Neural Networks and an application to Hollywood movies 483-487. [Crossref]

1670. Martin Wöllmer, Felix Weninger, Jürgen Geiger, Björn Schuller, Gerhard Rigoll. 2013. Noise robust ASR in reverberated multisource environments applying convolutive NMF and Long Short-Term Memory. *Computer Speech & Language* **27**:3, 780-797. [Crossref]

1671. Nicolas Boulanger-Lewandowski, Yoshua Bengio, Pascal Vincent. High-dimensional sequence transduction 3178-3182. [Crossref]

1672. Martin Wollmer, Zixing Zhang, Felix Weninger, Bjorn Schuller, Gerhard Rigoll. Feature enhancement by bidirectional LSTM networks for conversational speech recognition in highly non-stationary noise 6822-6826. [Crossref]

1673. DEREK MONNER, KAREN VATZ, GIOVANNA MORINI, SO-ONE HWANG, ROBERT DeKEYSER. 2013. A neural network model of the effects of entrenchment and memory development on grammatical gender learning. *Bilingualism: Language and Cognition* **16**:02, 246-265. [Crossref]

1674. Adam Eck, Leen-Kiat Soh. 2013. Observer effect from stateful resources in agent sensing. *Autonomous Agents and Multi-Agent Systems* **26**:2, 202-244. [Crossref]

1675. A. Alalshekmubarak, L. S. Smith. A novel approach combining recurrent neural network and support vector machines for time series classification 42-47. [Crossref]

1676. Martin Wöllmer, Björn Schuller, Gerhard Rigoll. 2013. Keyword spotting exploiting Long Short-Term Memory. *Speech Communication* **55**:2, 252-265. [Crossref]

1677. Angeliki Metallinou, Athanasios Katsamanis, Shrikanth Narayanan. 2013. Tracking continuous emotional trends of participants during affective dyadic interactions using body language and speech information. *Image and Vision Computing* **31**:2, 137-152. [Crossref]

1678. Martin Wöllmer, Moritz Kaiser, Florian Eyben, Björn Schuller, Gerhard Rigoll. 2013. LSTM-Modeling of continuous emotions in an audiovisual affect recognition framework. *Image and Vision Computing* **31**:2, 153-163. [Crossref]

1679. Esam Qaralleh, Gheith Abandah, Fuad Jamour. 2013. Tuning Recurrent Neural Networks for Recognizing Handwritten Arabic Words. *Journal of Software Engineering and Applications* **06**:10, 533-542. [Crossref]

1680. D. Badre. 2012. Opening the gate to working memory. *Proceedings of the National Academy of Sciences* **109**:49, 19878-19879. [Crossref]

1681. K. D'Ardenne, N. Eshel, J. Luka, A. Lenartowicz, L. E. Nystrom, J. D. Cohen. 2012. Role of prefrontal cortex and the midbrain dopamine system in working memory updating. *Proceedings of the National Academy of Sciences* **109**:49, 19900-19909. [Crossref]

1682. David C. Noelle. 2012. On the neural basis of rule-guided behavior. *Journal of Integrative Neuroscience* **11**:04, 453-475. [Crossref]

1683. Peter Dayan. 2012. How to set the switches on this thing. *Current Opinion in Neurobiology* **22**:6, 1068-1074. [Crossref]

1684. Derek Monner, James A. Reggia. 2012. Emergent latent symbol systems in recurrent neural networks. *Connection Science* **24**:4, 193-225. [Crossref]

1685. Emanuele Principi, Rudy Rotili, Martin Wöllmer, Florian Eyben, Stefano Squartini, Björn Schuller. 2012. Real-Time Activity Detection in a Multi-Talker Reverberated Environment. *Cognitive Computation* **4**:4, 386-397. [Crossref]

1686. Tomas Mikolov, Geoffrey Zweig. Context dependent recurrent neural network language model 234-239. [Crossref]

1687. Mantas Lukoševičius, Herbert Jaeger, Benjamin Schrauwen. 2012. Reservoir Computing Trends. *KI - Künstliche Intelligenz* **26**:4, 365-371. [Crossref]

1688. Barbara Hammer. 2012. Challenges in Neural Computation. *KI - Künstliche Intelligenz* **26**:4, 333-340. [Crossref]

1689. T. Waegeman, Francis Wyffels, B. Schrauwen. 2012. Feedback Control by Online Learning an Inverse Model. *IEEE Transactions on Neural Networks and Learning Systems* **23**:10, 1637-1648. [Crossref]

1690. Byron Leite Dantas Bezerra, Cleber Zanchettin, Vinicius Braga de Andrade. A Hybrid RNN Model for Cursive Offline Handwriting Recognition 113-118. [Crossref]

1691. Derek Monner, James A. Reggia. 2012. Neural architectures for learning to answer questions. *Biologically Inspired Cognitive Architectures* **2**, 37-53. [Crossref]

1692. Sebastian Otte, Dirk Krechel, Marcus Liwicki, Andreas Dengel. Local Feature Based Online Mode Detection with Recurrent Neural Networks 533-537. [Crossref]

1693. Jason M. Scimeca, David Badre. 2012. Striatal Contributions to Declarative Memory Retrieval. *Neuron* **75**:3, 380-392. [Crossref]

1694. Florin Schimbinschi, Marco Wiering, Rajesh Elara Mohan, Jaichandar Kulandaidaasan Sheba. 4D unconstrained real-time face recognition using a commodity depth camera 166-173. [Crossref]

1695. Vu-Anh Nguyen, Janusz A. Starzyk, Wooi-Boon Goh. Sequence recognition with spatio-temporal long-term memory organization 1-8. [Crossref]

1696. V. A. Nguyen, J. A. Starzyk, Wooi-Boon Goh, D. Jachyra. 2012. Neural Network Structure for Spatio-Temporal Long-Term Memory. *IEEE Transactions on Neural Networks and Learning Systems* **23**:6, 971-983. [Crossref]

1697. David Silver, Richard S. Sutton, Martin Müller. 2012. Temporal-difference search in computer Go. *Machine Learning* **87**:2, 183-219. [Crossref]

1698. A. Metallinou, M. Wollmer, A. Katsamanis, F. Eyben, B. Schuller, S. Narayanan. 2012. Context-Sensitive Learning for Enhanced Audiovisual Emotion Classification. *IEEE Transactions on Affective Computing* **3**:2, 184-198. [Crossref]

1699. Mounir Ben Nasr, Mohamed Chtourou. 2012. Training recurrent neural networks using a hybrid algorithm. *Neural Computing and Applications* **21**:3, 489-496. [Crossref]

1700. Juyang Weng. 2012. Symbolic Models and Emergent Models: A Review. *IEEE Transactions on Autonomous Mental Development* **4**:1, 29-53. [Crossref]

1701. Martin Wollmer, Angeliki Metallinou, Nassos Katsamanis, Bjorn Schuller, Shrikanth Narayanan. Analyzing the memory of BLSTM Neural Networks for enhanced emotion classification in dyadic spoken interactions 4157-4160. [Crossref]

1702. Felix Weninger, Martin Wollmer, Jurgen Geiger, Bjorn Schuller, Jort F. Gemmeke, Antti Hurmalainen, Tuomas Virtanen, Gerhard Rigoll. Non-negative matrix factorization for highly noise-robust ASR: To enhance or to recognize? 4681-4684. [Crossref]

1703. Sebastian Bock, Markus Schedl. Polyphonic piano note transcription with recurrent neural networks 121-124. [Crossref]

1704. Florian Eyben, Stavros Petridis, Bjorn Schuller, Maja Pantic. Audiovisual vocal outburst classification in noisy acoustic conditions 5097-5100. [Crossref]

1705. Michael J. Frank, David Badre. 2012. Mechanisms of Hierarchical Reinforcement Learning in Corticostriatal Circuits 1: Computational Analysis. *Cerebral Cortex* **22**:3, 509-526. [Crossref]

1706. Henning Schroll, Julien Vitay, Fred H. Hamker. 2012. Working memory and response selection: A computational account of interactions among cortico-basalganglio-thalamic loops. *Neural Networks* **26**, 59-74. [Crossref]

1707. James Martens, Ilya Sutskever. Training Deep and Recurrent Networks with Hessian-Free Optimization 479-535. [Crossref]

1708. Dan Claudiu Cireşan, Ueli Meier, Luca Maria Gambardella, Jürgen Schmidhuber. Deep Big Multilayer Perceptrons for Digit Recognition 581-598. [Crossref]

1709. Derek Monner, James A. Reggia. 2012. A generalized LSTM-like training algorithm for second-order recurrent neural networks. *Neural Networks* **25**, 70-83. [Crossref]

1710. Mantas Lukoševičius. A Practical Guide to Applying Echo State Networks 659-686. [Crossref]

1711. Andrey V. Gavrilov. 100. [Crossref]

1712. Laura Caponetti, Cosimo Alessandro Buscicchio, Giovanna Castellano. 2011. Biologically inspired emotion recognition from speech. *EURASIP Journal on Advances in Signal Processing* **2011**:1. . [Crossref]

1713. Martin Wollmer, Bjorn Schuller, Gerhard Rigoll. A novel bottleneck-BLSTM front-end for feature-level context modeling in conversational speech recognition 36-41. [Crossref]

1714. Karel Vesely, Martin Karafiat, Frantisek Grezl. Convolutive Bottleneck Network features for LVCSR 42-47. [Crossref]

1715. W. Heidl, C. Eitzinger, M. Gyimesi, F. Breitenecker. 2011. Learning over sets with Recurrent Neural Networks: An empirical categorization of aggregation functions. *Mathematics and Computers in Simulation* **82**:3, 442-449. [Crossref]

1716. François Rhéaume, Dominic Grenier, Éloi Bossé. 2011. Multistate combination approaches for liquid state machine in supervised spatiotemporal pattern classification. *Neurocomputing* **74**:17, 2842-2851. [Crossref]

1717. . Sequence Learning 190-216. [Crossref]

1718. Martin Wöllmer, Erik Marchi, Stefano Squartini, Björn Schuller. 2011. Multi-stream LSTM-HMM decoding and histogram equalization for noise robust keyword spotting. *Cognitive Neurodynamics* **5**:3, 253-264. [Crossref]

1719. Martin Wöllmer, Felix Weninger, Florian Eyben, Björn Schuller. 2011. Computational Assessment of Interest in Speech—Facing the Real-Life Challenge. *KI - Künstliche Intelligenz* **25**:3, 225-234. [Crossref]

1720. Renato Ramos da Silva, Roseli Aparecida Francelin Romero. Relational reinforcement learning and recurrent neural network with state classification to solve joint attention 1222-1229. [Crossref]

1721. Hyung O. Park, Alireza A. Dibazar, Theodore W. Berger. Discrete Synapse Recurrent Neural Network with time-varying delays for nonlinear system modeling and its application on seismic signal classification 2374-2381. [Crossref]

1722. Trent Kriete, David C. Noelle. 2011. Generalisation benefits of output gating in a model of prefrontal cortex. *Connection Science* **23**:2, 119-129. [Crossref]

1723. M Wollmer, C Blaschke, T Schindl, B Schuller, B Farber, S Mayer, B Trefflich. 2011. Online Driver Distraction Detection Using Long Short-Term Memory. *IEEE Transactions on Intelligent Transportation Systems* **12**:2, 574-582. [Crossref]

1724. Christian Landsiedel, Jens Edlund, Florian Eyben, Daniel Neiberg, Bjorn Schuller. Syllabification of conversational speech using Bidirectional Long-Short-Term Memory Neural Networks 5256-5259. [Crossref]

1725. Martin Wollmer, Florian Eyben, Bjorn Schuller, Gerhard Rigoll. A multi-stream ASR framework for BLSTM modeling of conversational speech 4860-4863. [Crossref]

1726. Stavros Petridis, Maja Pantic. 2011. Audiovisual Discrimination Between Speech and Laughter: Why and When Visual Information Might Help. *IEEE Transactions on Multimedia* **13**:2, 216-234. [Crossref]

1727. David Badre. 2011. Defining an Ontology of Cognitive Control Requires Attention to Component Interactions. *Topics in Cognitive Science* **3**:2, 217-221. [Crossref]

1728. Mounir Ben Nasr, M Chtourou. On the training of recurrent neural networks 1-5. [Crossref]

1729. Haikal El Abed, Volker Märgner. 2011. ICDAR 2009-Arabic handwriting recognition competition. *International Journal on Document Analysis and Recognition (IJDAR)* **14**:1, 3-13. [Crossref]

1730. Haikal El Abed, Monji Kherallah, Volker Märgner, Adel M. Alimi. 2011. On-line Arabic handwriting recognition competition. *International Journal on Document Analysis and Recognition (IJDAR)* **14**:1, 15-23. [Crossref]

1731. Martin Wöllmer, Erik Marchi, Stefano Squartini, Björn Schuller. Robust Multi-stream Keyword and Non-linguistic Vocalization Detection for Computationally Intelligent Virtual Agents 496-505. [Crossref]

1732. Thomas Rückstieß, Christian Osendorfer, Patrick van der Smagt. Sequential Feature Selection for Classification 132-141. [Crossref]

1733. Marcus Liwicki, Horst Bunke, James A. Pittman, Stefan Knerr. 2011. Combining diverse systems for handwritten text line recognition. *Machine Vision and Applications* **22**:1, 39-51. [Crossref]

1734. Martin Wöllmer, Björn Schuller. Enhancing Spontaneous Speech Recognition with BLSTM Features 17-24. [Crossref]

1735. Nikolay Nikolaev, Peter Tino, Evgueni Smirnov. Time-Dependent Series Variance Estimation via Recurrent Neural Networks 176-184. [Crossref]

1736. Paul D. Yoo, Bing Bing Zhou, Albert Y. Zomaya. Protein Domain Boundary Prediction 501-519. [Crossref]

1737. Dan Claudiu Cireşan, Ueli Meier, Luca Maria Gambardella, Jürgen Schmidhuber. 2010. Deep, Big, Simple Neural Nets for Handwritten Digit Recognition. *Neural Computation* **22**:12, 3207-3220. [Abstract] [Full Text] [PDF] [PDF Plus]

1738. B. Roeschies, C. Igel. 2010. Structure optimization of reservoir networks. *Logic Journal of IGPL* **18**:5, 635-669. [Crossref]

1739. Martin Wollmer, Björn Schuller, Florian Eyben, Gerhard Rigoll. 2010. Combining Long Short-Term Memory and Dynamic Bayesian Networks for Incremental Emotion-Sensitive Artificial Listening. *IEEE Journal of Selected Topics in Signal Processing* **4**:5, 867-881. [Crossref]

1740. D. Wierstra, A. Forster, J. Peters, J. Schmidhuber. 2010. Recurrent policy gradients. *Logic Journal of IGPL* **18**:5, 620-634. [Crossref]

1741. Martin Wöllmer, Florian Eyben, Alex Graves, Björn Schuller, Gerhard Rigoll. 2010. Bidirectional LSTM Networks for Context-Sensitive Keyword Detection in a Cognitive Virtual Agent Framework. *Cognitive Computation* **2**:3, 180-190. [Crossref]

1742. Jürgen Schmidhuber. 2010. Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development* **2**:3, 230-247. [Crossref]

1743. Sanjay Jain, Steffen Lange, Samuel E. Moelius, Sandra Zilles. 2010. Incremental learning with temporary memory. *Theoretical Computer Science* **411**:29-30, 2757-2772. [Crossref]

1744. Jun Namikawa, Jun Tani. 2010. Learning to imitate stochastic time series in a compositional way by chaos. *Neural Networks* **23**:5, 625-638. [Crossref]

1745. David Panzoli, Sara de Freitas, Yves Duthen, Hervé Luga. 2010. The Cortexionist architecture: behavioural intelligence of artificial creatures. *The Visual Computer* **26**:5, 353-366. [Crossref]

1746. Randall C O'Reilly, Seth A Herd, Wolfgang M Pauli. 2010. Computational models of cognitive control. *Current Opinion in Neurobiology* **20**:2, 257-261. [Crossref]

1747. Florian Eyben, Martin Wöllmer, Alex Graves, Björn Schuller, Ellen Douglas-Cowie, Roddy Cowie. 2010. On-line emotion recognition in a 3-D activation-valence-time continuum using acoustic and linguistic cues. *Journal on Multimodal User Interfaces* **3**:1-2, 7-19. [Crossref]

1748. Ilya Sutskever, Geoffrey Hinton. 2010. Temporal-Kernel Recurrent Neural Networks. *Neural Networks* **23**:2, 239-243. [Crossref]

1749. Thomas E. Portegys. 2010. A maze learning comparison of Elman, long short-term memory, and Mona neural networks. *Neural Networks* **23**:2, 306-313. [Crossref]

1750. François Rivest, John F. Kalaska, Yoshua Bengio. 2010. Alternative time representation in dopamine models. *Journal of Computational Neuroscience* **28**:1, 107-130. [Crossref]

1751. Florian Eyben, Martin Wöllmer, Tony Poitschke, Björn Schuller, Christoph Blaschke, Berthold Färber, Nhu Nguyen-Thien. 2010. Emotion on the Road—Necessity, Acceptance, and Feasibility of Affective Computing in the Car. *Advances in Human-Computer Interaction* **2010**, 1-17. [Crossref]

1752. Martin Wollmer, Florian Eyben, Bjorn Schuller, Gerhard Rigoll. Spoken term detection with Connectionist Temporal Classification: A novel hybrid CTC-DBN decoder 5274-5277. [Crossref]

1753. Bjorn Schuller, Felix Weninger, Martin Wollmer, Yang Sun, Gerhard Rigoll. Non-negative matrix factorization as noise-robust feature extractor for speech recognition 4562-4565. [Crossref]

1754. Peng Zhao, O. P. Malik. 2009. Design of an Adaptive PSS Based on Recurrent Adaptive Control Theory. *IEEE Transactions on Energy Conversion* **24**:4, 884-892. [Crossref]

1755. PETER STUBBERUD. 2009. A VECTOR MATRIX REAL TIME RECURSIVE BACKPROPAGATION ALGORITHM FOR RECURRENT NEURAL NETWORKS THAT APPROXIMATE MULTI-VALUED PERIODIC FUNCTIONS. *International Journal of Computational Intelligence and Applications* **08**:04, 395-411. [Crossref]

1756. Martin Wöllmer, Marc Al-Hames, Florian Eyben, Björn Schuller, Gerhard Rigoll. 2009. A multidimensional dynamic time warping algorithm for efficient multimodal fusion of asynchronous data streams. *Neurocomputing* **73**:1-3, 366-380. [Crossref]

1757. Joschka Boedecker, Oliver Obst, N. Michael Mayer, Minoru Asada. 2009. Initialization and self-organized optimization of recurrent neural network connectivity. *HFSP Journal* **3**:5, 340-349. [Crossref]

1758. Florian Eyben, Martin Wollmer, Bjorn Schuller. OpenEAR — Introducing the munich open-source emotion and affect recognition toolkit 1-6. [Crossref]

1759. Martin V. Butz, Thies D. Lonneker. Optimized sensory-motor couplings plus strategy extensions for the TORCS car racing challenge 317-324. [Crossref]

1760. E. Fuchs, C. Gruber, T. Reitmaier, B. Sick. 2009. Processing Short-Term and Long-Term Information With a Combination of Polynomial Approximation Techniques and Time-Delay Neural Networks. *IEEE Transactions on Neural Networks* **20**:9, 1450-1462. [Crossref]

1761. Amaury Hazan, Ricard Marxer, Paul Brossier, Hendrik Purwins, Perfecto Herrera, Xavier Serra. 2009. What/when causal expectation modelling applied to audio signals. *Connection Science* **21**:2-3, 119-143. [Crossref]

1762. Stefan J. Kiebel, Katharina von Kriegstein, Jean Daunizeau, Karl J. Friston. 2009. Recognizing Sequences of Sequences. *PLoS Computational Biology* **5**:8, e1000464. [Crossref]

1763. Mantas Lukoševičius, Herbert Jaeger. 2009. Reservoir computing approaches to recurrent neural network training. *Computer Science Review* **3**:3, 127-149. [Crossref]

1764. MARCUS LIWICKI, HORST BUNKE. 2009. FEATURE SELECTION FOR HMM AND BLSTM BASED HANDWRITING RECOGNITION OF WHITEBOARD NOTES. *International Journal of Pattern Recognition and Artificial Intelligence* **23**:05, 907-923. [Crossref]

1765. Barak A. Pearlmutter, Conor J. Houghton. 2009. A New Hypothesis for Sleep: Tuning for Criticality. *Neural Computation* **21**:6, 1622-1641. [Abstract] [Full Text] [PDF] [PDF Plus]

1766. Juyang Weng, Yi Shen, Mingmin Chi, Xiangyang Xue. Temporal context as cortical spatial codes 3348-3355. [Crossref]

1767. Gerry T. M. Altmann, Jelena Mirković. 2009. Incrementality and Prediction in Human Sentence Processing. *Cognitive Science* **33**:4, 583-609. [Crossref]

1768. A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, J. Schmidhuber. 2009. A Novel Connectionist System for Unconstrained Handwriting Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31**:5, 855-868. [Crossref]

1769. J.A. Starzyk, H. He. 2009. Spatio–Temporal Memories for Machine Learning: A Long-Term Memory Organization. *IEEE Transactions on Neural Networks* **20**:5, 768-780. [Crossref]

1770. Martin Wollmer, Florian Eyben, Joseph Keshet, Alex Graves, Bjorn Schuller, Gerhard Rigoll. Robust discriminative keyword spotting for emotionally colored spontaneous speech using bidirectional LSTM networks 3949-3952. [Crossref]

1771. J. Horn, O. De Jesus, M.T. Hagan. 2009. Spurious Valleys in the Error Surface of Recurrent Networks—Analysis and Avoidance. *IEEE Transactions on Neural Networks* **20**:4, 686-700. [Crossref]

1772. David Montana, Eric VanWyk, Marshall Brinn, Joshua Montana, Stephen Milligan. 2009. Evolution of internal dynamics for neural network nodes. *Evolutionary Intelligence* **1**:4, 233-251. [Crossref]

1773. Kai A. Krueger, Peter Dayan. 2009. Flexible shaping: How learning in small steps helps. *Cognition* **110**:3, 380-394. [Crossref]

1774. Ueli Rutishauser, Rodney J. Douglas. 2009. State-Dependent Computation Using Coupled Recurrent Networks. *Neural Computation* **21**:2, 478-509. [Abstract] [Full Text] [PDF] [PDF Plus]

1775. Andrew R.A. Conway, Adam B. Moore, Michael J. Kane. 2009. Recent trends in the cognitive neuroscience of working memory. *Cortex* **45**:2, 262-268. [Crossref]

1776. Justin Bayer, Daan Wierstra, Julian Togelius, Jürgen Schmidhuber. Evolving Memory Cell Structures for Sequence Learning 755-764. [Crossref]

1777. Haikal El Abed, Volker Märgner, Monji Kherallah, Adel M. Alimi. ICDAR 2009 Online Arabic Handwriting Recognition Competition 1388-1392. [Crossref]

1778. Emmanuèle Grosicki, Haikal El Abed. ICDAR 2009 Handwriting Recognition Competition 1398-1402. [Crossref]

1779. Wu Wei, Gao Guanglai. Online Handwriting Mongolia Words Recognition with Recurrent Neural Networks 165-167. [Crossref]

1780. David Panzoli, Hervé Luga, Yves Duthen. Beyond Reactive Systems: An Associative Memory for Sensory-Driven Intelligent Behavior 63-70. [Crossref]

1781. Tom Schaul, Jürgen Schmidhuber. Scalable Neural Networks for Board Games 1005-1014. [Crossref]

1782. Jürgen Schmidhuber. Driven by Compression Progress: A Simple Principle Explains Essential Aspects of Subjective Beauty, Novelty, Surprise, Interestingness, Attention, Curiosity, Creativity, Art, Science, Music, Jokes 48-76. [Crossref]

1783. Juyang Weng, Qi Zhang, Mingmin Chi, Xiangyang Xue. Complex text processing by the temporal context machines 1-8. [Crossref]

1784. Mohamad Faizal bin Samsudin, Katsunari Shibata. Improvement of Practical Recurrent Learning Method and Application to a Pattern Classification Task 631-638. [Crossref]

1785. Björn Schuller, Martin Wöllmer, Tobias Moosmayr, Gerhard Rigoll. 2009. Recognition of Noisy Speech: A Comparative Survey of Robust Model Architecture and Feature Enhancement. *EURASIP Journal on Audio, Speech, and Music Processing* **2009**:1, 942617. [Crossref]

1786. Michal Čerňanský, Ľubica Beňušková. Training Recurrent Connectionist Models on Symbolic Time Series 285-292. [Crossref]

1787. Jun Namikawa, Jun Tani. 2008. A model for learning to segment temporal sequences, utilizing a mixture of RNN experts together with adaptive variance. *Neural Networks* **21**:10, 1466-1475. [Crossref]

1788. Tom Schaul, Jurgen Schmidhuber. A scalable neural network architecture for board games 357-364. [Crossref]

1789. Fathy A. Elmisery, Janusz A. Starzyk. A neural network based on sequence learning for speech recognition 139-142. [Crossref]

1790. Brad Aisa, Brian Mingus, Randy O'Reilly. 2008. The Emergent neural modeling system. *Neural Networks* **21**:8, 1146-1152. [Crossref]

1791. Le Tien Dung, Takashi Komeda, Motoki Takagi. 2008. REINFORCEMENT LEARNING FOR POMDP USING STATE CLASSIFICATION. *Applied Artificial Intelligence* **22**:7-8, 761-779. [Crossref]

1792. Anton Maximilian Schaefer, Steffen Udluft, Hans-Georg Zimmermann. 2008. Learning long-term dependencies with recurrent neural networks. *Neurocomputing* **71**:13-15, 2481-2488. [Crossref]

1793. D. Novak, K. Mucha, T. Al-Ani. Long Short-Term Memory for apnea detection based on Heart Rate Variability 5234-5237. [Crossref]

1794. Le Tien Dung, Takashi Komeda, Motoki Takagi. Efficient experience reuse in non-Markovian environments 3327-3332. [Crossref]

1795. Débora C. Corrêa, Alexandre L. M. Levada, José Hiroki Saito. Stabilizing and Improving the Learning Speed of 2-Layered LSTM Network 293-300. [Crossref]

1796. Débora Cristina Corrêa, José Hiroki Saito, Sandra Abib. Composing Music with BPTT and LSTM Networks: Comparing Learning and Generalization Aspects 95-100. [Crossref]

1797. Eduardo Bayro-Corrochano, J. Refugio Vallejo-Gutierrez, Nancy Arana-Daniel. Recurrent Clifford Support Machines 3613-3618. [Crossref]

1798. Alexandre L. M. Levada, Debora C. Correa, Denis H. P. Salvadeo, Jose H. Saito, Nelson D. A. Mascarenhas. Novel approaches for face recognition: Template-matching using Dynamic Time Warping and LSTM neural network supervised classification 241-244. [Crossref]

1799. Benjamin Schrauwen, Marion Wardermann, David Verstraeten, Jochen J. Steil, Dirk Stroobandt. 2008. Improving reservoirs using intrinsic plasticity. *Neurocomputing* **71**:7-9, 1159-1171. [Crossref]

1800. Bo Cartling. 2008. On the implicit acquisition of a context-free grammar by a simple recurrent neural network. *Neurocomputing* **71**:7-9, 1527-1537. [Crossref]

1801. Hermann Mayer, Faustino Gomez, Daan Wierstra, Istvan Nagy, Alois Knoll, Jürgen Schmidhuber. 2008. A System for Robotic Heart Surgery that Learns to Tie Knots Using Recurrent Neural Networks. *Advanced Robotics* **22**:13-14, 1521-1537. [Crossref]

1802. F. Cadini, E. Zio, N. Pedroni. 2008. Validation of Infinite Impulse Response Multilayer Perceptron for Modelling Nuclear Dynamics. *Science and Technology of Nuclear Installations* **2008**, 1-10. [Crossref]

1803. André Grüning. 2007. Elman Backpropagation as Reinforcement for Simple Recurrent Networks. *Neural Computation* **19**:11, 3108-3131. [Abstract] [PDF] [PDF Plus]

1804. Le Tien Dung, Takashi Komeda, Motoki Takagi. Reinforcement learning in non-markovian environments using automatic discovery of subgoals 2601-2605. [Crossref]

1805. Garen Arevian, Christo Panchev. Optimising the Hystereses of a Two Context Layer RNN for Text Classification 2936-2941. [Crossref]

1806. S. Hochreiter, M. Heusel, K. Obermayer. 2007. Fast model-based protein homology detection without alignment. *Bioinformatics* **23**:14, 1728-1736. [Crossref]

1807. Jeremy R. Reynolds, Jeffrey M. Zacks, Todd S. Braver. 2007. A Computational Model of Event Segmentation From Perceptual Prediction. *Cognitive Science* **31**:4, 613-643. [Crossref]

1808. T. Thireou, M. Reczko. 2007. Bidirectional Long Short-Term Memory Networks for Predicting the Subcellular Localization of Eukaryotic Proteins. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **4**:3, 441-446. [Crossref]

1809. Le Tien Dung, Takashi Komeda, Motoki Takagi. Mixed Reinforcement Learning for Partially Observable Markov Decision Process 7-12. [Crossref]

1810. F. Cadini, E. Zio, N. Pedroni. 2007. Simulating the dynamics of the neutron flux in a nuclear reactor by locally recurrent neural networks. *Annals of Nuclear Energy* **34**:6, 483-495. [Crossref]

1811. P.R. Davidson, R.D. Jones, M.T.R. Peiris. 2007. EEG-Based Lapse Detection With High Temporal Resolution. *IEEE Transactions on Biomedical Engineering* **54**:5, 832-839. [Crossref]

1812. Anton Maximilian Schaefer, Steffen Udluft, Hans-Georg Zimmermann. A Recurrent Control Neural Network for Data Efficient Reinforcement Learning 151-157. [Crossref]

1813. Bram Bakker. Reinforcement learning by backpropagation through an LSTM model/critic 127-134. [Crossref]

1814. Jürgen Schmidhuber, Daan Wierstra, Matteo Gagliolo, Faustino Gomez. 2007. Training Recurrent Networks by Evolino. *Neural Computation* **19**:3, 757-779. [Abstract] [PDF] [PDF Plus]

1815. Janusz A. Starzyk, Haibo He. 2007. Anticipation-Based Temporal Sequences Learning in Hierarchical Structure. *IEEE Transactions on Neural Networks* **18**:2, 344-358. [Crossref]

1816. Mustafa C. Ozturk, Dongming Xu, José C. Príncipe. 2007. Analysis and Design of Echo State Networks. *Neural Computation* **19**:1, 111-138. [Abstract] [PDF] [PDF Plus]

1817. Orlando De Jesus, M.T. Hagan. 2007. Backpropagation Algorithms for a Broad Class of Dynamic Networks. *IEEE Transactions on Neural Networks* **18**:1, 14-27. [Crossref]

1818. Hermann Mayer, Faustino Gomez, Daan Wierstra, Istvan Nagy, Alois Knoll, Jurgen Schmidhuber. A System for Robotic Heart Surgery that Learns to Tie Knots Using Recurrent Neural Networks 543-548. [Crossref]

1819. Judy A. Franklin. 2006. Recurrent Neural Networks for Music Computation. *INFORMS Journal on Computing* **18**:3, 321-338. [Crossref]

1820. JUDY A. FRANKLIN. 2006. JAZZ MELODY GENERATION USING RECURRENT NETWORKS AND REINFORCEMENT LEARNING. *International Journal on Artificial Intelligence Tools* **15**:04, 623-650. [Crossref]

1821. Jürgen Schmidhuber. 2006. Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connection Science* **18**:2, 173-187. [Crossref]

1822. André Grüning. 2006. Stack-like and queue-like dynamics in recurrent neural networks. *Connection Science* **18**:1, 23-42. [Crossref]

1823. Randall C. O'Reilly, Michael J. Frank. 2006. Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia. *Neural Computation* **18**:2, 283-328. [Abstract] [PDF] [PDF Plus]

1824. L.E. Hibell, Honghai Liu, D.J. Brown. Combining Multi-Frame Images for Enhancement Using Self-Delaying Dynamic Networks 4817-4824. [Crossref]

1825. Matthew M. Botvinick, David C. Plaut. 2006. Short-term memory for serial order: A recurrent neural network model. *Psychological Review* **113**:2, 201-233. [Crossref]

1826. Ali Zifan, Shahriar Gharibzadeh. 2006. The chaotic nature of temper in humans: A long short-term memory recurrent neural network model. *Medical Hypotheses* **67**:3, 658-661. [Crossref]

1827. Alessio Ceroni, Paolo Frasconi, Gianluca Pollastri. 2005. Learning protein secondary structure from sequential and relational data. *Neural Networks* **18**:8, 1029-1039. [Crossref]

1828. Thomas Voegtlin. 2005. Recursive principal components analysis. *Neural Networks* **18**:8, 1051-1063. [Crossref]

1829. J.W.C. van Lint, S.P. Hoogendoorn, H.J. van Zuylen. 2005. Accurate freeway travel time prediction with state-space neural networks under missing data. *Transportation Research Part C: Emerging Technologies* **13**:5-6, 347-369. [Crossref]

1830. M. Delgado, M.C. Pegalajar. 2005. A multiobjective genetic algorithm for obtaining the optimal size of a recurrent neural network for grammatical inference. *Pattern Recognition* **38**:9, 1444-1456. [Crossref]

1831. Alex Graves, Jürgen Schmidhuber. 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks* **18**:5-6, 602-610. [Crossref]

1832. Y. Barniv, M. Aguilar, E. Hasanbelliu. 2005. Using EMG to Anticipate Head Motion for Virtual-Environment Applications. *IEEE Transactions on Biomedical Engineering* **52**:6, 1078-1093. [Crossref]

1833. N. P. Rougier, D. C. Noelle, T. S. Braver, J. D. Cohen, R. C. O'Reilly. 2005. Prefrontal cortex and flexible cognitive control: Rules without symbols. *Proceedings of the National Academy of Sciences* **102**:20, 7338-7343. [Crossref]

1834. M. Boden, J. Hawkins. 2005. Improved Access to Sequential Motifs: A Note on the Architectural Bias of Recurrent Networks. *IEEE Transactions on Neural Networks* **16**:2, 491-494. [Crossref]

1835. Siu-Yeung Cho, Zheru Chi. 2005. Genetic evolution processing of data structures for image classification. *IEEE Transactions on Knowledge and Data Engineering* **17**:2, 216-231. [Crossref]

1836. JUDY A. FRANKLIN, KRYSTAL K. LOCKE. 2005. RECURRENT NEURAL NETWORKS FOR MUSICAL PITCH MEMORY AND CLASSIFICATION. *International Journal on Artificial Intelligence Tools* **14**:01n02, 329-342. [Crossref]

1837. A.G. Tijsseling. 2005. Sequential Information Processing Using Time-Delay Connections in Ontogenic CALM Networks. *IEEE Transactions on Neural Networks* **16**:1, 145-159. [Crossref]

1838. Yanjun Shen, Bingwen Wang, Fangxin Chen, Liang Cheng. 2004. A New Multi-output Neural Model with Tunable Activation Function and its Applications. *Neural Processing Letters* **20**:2, 85-104. [Crossref]

1839. Rainer W. Paine, Jun Tani. 2004. Motor primitive and sequence self-organization in a hierarchical recurrent neural network. *Neural Networks* **17**:8-9, 1291-1309. [Crossref]

1840. Barbara Hammer, Alessio Micheli, Alessandro Sperduti, Marc Strickert. 2004. A general framework for unsupervised processing of structured data. *Neurocomputing* **57**, 3-35. [Crossref]

1841. P. Tino, M. Cernansky, L. Benuskova. 2004. Markovian Architectural Bias of Recurrent Neural Networks. *IEEE Transactions on Neural Networks* **15**:1, 6-15. [Crossref]

1842. Barbara Hammer, Peter Tiňo. 2003. Recurrent Neural Networks with Small Weights Implement Definite Memory Machines. *Neural Computation* **15**:8, 1897-1929. [Abstract] [PDF] [PDF Plus]

1843. Siu-Yeung Cho, Zheru Chi, Wan-Chi Siu, Ah Chung Tsoi. 2003. An improved algorithm for learning long-term dependency problems in adaptive processing of data structures. *IEEE Transactions on Neural Networks* **14**:4, 781-793. [Crossref]

1844. Guilherme de A. Barreto, Aluizio F. R. Araújo, Stefan C. Kremer. 2003. A Taxonomy for Spatiotemporal Connectionist Networks Revisited: The Unsupervised Case. *Neural Computation* **15**:6, 1255-1320. [Abstract] [PDF] [PDF Plus]

1845. Juan Antonio Pérez-Ortiz, Felix A. Gers, Douglas Eck, Jürgen Schmidhuber. 2003. Kalman filters improve LSTM network performance in problems unsolvable by traditional recurrent nets. *Neural Networks* **16**:2, 241-250. [Crossref]

1846. Ivan Gabrijel, Andrej Dobnikar. 2003. On-line identification and reconstruction of finite automata with generalized recurrent neural networks. *Neural Networks* **16**:1, 101-120. [Crossref]

1847. Jun Tani. 2003. Learning to generate articulated behavior through the bottom-up and the top-down interaction processes. *Neural Networks* **16**:1, 11-23. [Crossref]

1848. Andrew D. Back, Tianping Chen. 2002. Universal Approximation of Multiple Nonlinear Operators by Neural Networks. *Neural Computation* **14**:11, 2561-2566. [Abstract] [PDF] [PDF Plus]

1849. J. Schmidhuber, F. Gers, D. Eck. 2002. Learning Nonregular Languages: A Comparison of Simple Recurrent Networks and LSTM. *Neural Computation* **14**:9, 2039-2041. [Abstract] [PDF] [PDF Plus]

1850. Alex Aussem. 2002. Sufficient Conditions for Error Backflow Convergence in Dynamical Recurrent Neural Networks. *Neural Computation* **14**:8, 1907-1927. [Abstract] [PDF] [PDF Plus]

1851. Jonathan D Cohen, Todd S Braver, Joshua W Brown. 2002. Computational perspectives on dopamine function in prefrontal cortex. *Current Opinion in Neurobiology* **12**:2, 223-229. [Crossref]

1852. F.A. Gers, E. Schmidhuber. 2001. LSTM recurrent networks learn simple context-free and context-sensitive languages. *IEEE Transactions on Neural Networks* **12**:6, 1333-1340. [Crossref]

1853. P.C.R. Lane, J.B. Henderson. 2001. Incremental syntactic parsing of natural language corpora with simple synchrony networks. *IEEE Transactions on Knowledge and Data Engineering* **13**:2, 219-231. [Crossref]

1854. B. Hammer. 2001. Generalization ability of folding networks. *IEEE Transactions on Knowledge and Data Engineering* **13**:2, 196-206. [Crossref]

1855. H.R. Maier, G.C. Dandy. 2001. Neural network based modelling of environmental variables: A systematic approach. *Mathematical and Computer Modelling* **33**:6-7, 669-682. [Crossref]

1856. Stefan C. Kremer. 2001. Spatiotemporal Connectionist Networks: A Taxonomy and Review. *Neural Computation* **13**:2, 249-306. [Abstract] [PDF] [PDF Plus]

1857. Felix A. Gers, Jürgen Schmidhuber, Fred Cummins. 2000. Learning to Forget: Continual Prediction with LSTM. *Neural Computation* **12**:10, 2451-2471. [Abstract] [PDF] [PDF Plus]

1858. Jyh-Da Wei, Chuen-Tsai Sun. 2000. Constructing hysteretic memory in neural networks. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)* **30**:4, 601-609. [Crossref]

1859. A.F. Atiya, A.G. Parlos. 2000. New results on recurrent network training: unifying the algorithms and accelerating convergence. *IEEE Transactions on Neural Networks* **11**:3, 697-709. [Crossref]

1860. D.P. Mandic, J.A. Chambers. 2000. On the choice of parameters of the cost function in nested modular RNN's. *IEEE Transactions on Neural Networks* **11**:2, 315-322. [Crossref]

1861. Barbara Hammer. 2000. On the approximation capability of recurrent neural networks. *Neurocomputing* **31**:1-4, 107-123. [Crossref]

1862. F.A. Gers, J. Schmidhuber. Recurrent nets that time and count 189-194 vol.3. [Crossref]

1863. A. Aussem. Sufficient conditions for error back flow convergence in dynamical recurrent neural networks 577-582 vol.4. [Crossref]

1864. F.A. Gers, J. Schmidhuber. Neural processing of complex continual input streams 557-562 vol.4. [Crossref]

1865. Holger R. Maier, Graeme C. Dandy. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software* **15**:1, 101-124. [Crossref]

1866. J. Tani, S. Nolfi. 1999. Learning to perceive the world as articulated: an approach for hierarchical learning in sensory-motor systems. *Neural Networks* **12**:7-8, 1131-1141. [Crossref]

1867. Todd S Braver, Deanna M Barch, Jonathan D Cohen. 1999. Cognition and control in schizophrenia: a computational model of dopamine and prefrontal function. *Biological Psychiatry* **46**:3, 312-328. [Crossref]

1868. P. Campolucci, A. Uncini, F. Piazza, B.D. Rao. 1999. On-line learning algorithms for locally recurrent neural networks. *IEEE Transactions on Neural Networks* **10**:2, 253-271. [Crossref]

1869. Ralph E. Hoffman, Thomas H. McGlashan. Chapter 18 Using a speech perception neural network simulation to explore normal neurodevelopment and hallucinated 'voices' in schizophrenia 311-325. [Crossref]

1870. Todd S. Braver, Jonathan D. Cohen. Chapter 19 Dopamine, cognitive control, and schizophrenia: the gating model 327-349. [Crossref]

1871. Tsungnan Lin, Bill G. Horne, C.Lee Giles. 1998. How embedded memory in recurrent neural network architectures helps learning long-term temporal dependencies. *Neural Networks* **11**:5, 861-868. [Crossref]

1872. Marco Wiering, Jürgen Schmidhuber. 1997. HQ-Learning. *Adaptive Behavior* **6**:2, 219-246. [Crossref]

1873. Tsungnan Lin, B.G. Horne, P. Tino, C.L. Giles. 1996. Learning long-term dependencies in NARX recurrent neural networks. *IEEE Transactions on Neural Networks* **7**:6, 1329-1338. [Crossref]

1874. B. Bakker, V. Zhumatiy, G. Gruener, J. Schmidhuber. A robot that reinforcement-learns to identify and memorize important previous observations 430-435. [Crossref]

1875. A.S. Younger, S. Hochreiter, P.R. Conwell. Meta-learning with backpropagation 2001-2006. [Crossref]

1876. M. Oubbati, P. Levi, M. Schanz. Meta-Learning for Adaptive Identification of Non-Linear Dynamical Systems 473-478. [Crossref]

1877. J. Bose, S.B. Furber, J.L. Shapiro. An associative memory for the on-line recognition and prediction of temporal sequences 1223-1228. [Crossref]

1878. Leonard Johard, Vittorio Lippi, Larisa Safina, Manuel Mazzara. Mind and Matter: 63-82. [Crossref]

1879. A. Graves, J. Schmidhuber. Framewise phoneme classification with bidirectional LSTM networks 2047-2052. [Crossref]

1880. Gee Wah Ng, Chung Huat Tan, N. Chaturvedi. Building intelligent system: learning from the Cortico-Hippocampal network 1248-1253. [Crossref]

1881. J.J. Steil. Backpropagation-decorrelation: online recurrent learning with O(N) complexity 843-848. [Crossref]

1882. A.M. Fanelli, L. Caponetti, G. Castellano, C.A. Buscicchio. Content-based recognition of musical instruments 361-364. [Crossref]

1883. N. Bergfeldt, F. Linaker. Self-organized modulation of a neural robot controller 495-500. [Crossref]

1884. R.A. Santiago. Context discerning multifunction networks: reformulating fixed weight neural networks 189-194. [Crossref]

1885. S. Squartini, S. Paolinelli, F. Piazza. Comparing Different Recurrent Neural Architectures on a Specific Task from Vanishing Gradient Effect Perspective 380-385. [Crossref]

1886. I. Szita, A. Lorincz. Simple algorithm for recurrent neural networks that can learn sequence completion 183-188. [Crossref]

1887. S. Reiter, B. Schuller, G. Rigoll. A Combined LSTM-RNN - HMM - Approach for Meeting Event Segmentation and Recognition II-393-II-396. [Crossref]

1888. S. Hochreiter, K. Obermayer. Optimal gradient-based learning using importance weights 114-119. [Crossref]