Visualizing and Understanding Recurrent Neural Networks

Andrej Karpathy*, Justin Johnson*, Fei-Fei Li
Automated Image Captioning with ConvNets and Recurrent Nets

Andrej Karpathy, Fei-Fei Li
man in black shirt is playing guitar.

construction worker in orange safety vest is working on road.

two young girls are playing with lego toy.

boy is doing backflip on wakeboard.

girl in pink dress is jumping in air.

black and white dog jumps over bar.

young girl in pink shirt is swinging on swing.

man in blue wetsuit is surfing on wave.
Describing images

Recurrent Neural Network

Convolutional Neural Network
Recurrent Neural Network
Recurrent Networks offer a lot of flexibility:
Sequential Processing

Multiple Object Recognition with Visual Attention, Ba et al.
Language Models

Word-level language model. Similar to:

Recurrent Neural Network Based Language Model
[Tomas Mikolov, 2010]
test image

sample!
test image
test image

sample!

<START>
test image
test image

sample! <END> token => finish.
Image Sentence Datasets

Microsoft COCO
[Tsung-Yi Lin et al. 2014]
mscoco.org

currently:
~120K images
~5 sentences each
Example Results
Wow I can’t believe that worked
Wow I can’t believe that worked

- A cat is sitting on a toilet seat
  logprob: -7.79

- A display case filled with lots of different types of donuts
  logprob: -7.78

- A group of people sitting at a table with wine glasses
  logprob: -6.71
Well, I can kind of see it
Well, I can kind of see it
Not sure what happened there...

- a toilet with a seat up in a bathroom  
  logprob: -13.44
- a woman holding a teddy bear in front of a mirror  
  logprob: -9.65
- a horse is standing in the middle of a road  
  logprob: -10.34
Course Description

Computer Vision has become ubiquitous in our society, with applications in search, image understanding, apps, mapping, medicine, drones, and self-driving cars. Core to many of these applications are visual recognition tasks such as image classification, localization and detection. Recent developments in neural network (aka “deep learning”) approaches have greatly advanced the performance of these state-of-the-art visual recognition systems. This course is a deep dive into details of the deep learning architectures with a focus on learning end-to-end models for these tasks, particularly image classification. During the 10-week course, students will learn to implement, train and debug their own neural networks and gain a detailed understanding of cutting-edge research in computer vision. The final assignment will involve training a multi-million parameter convolutional neural network and applying it on the largest image classification dataset (ImageNet). We will focus on teaching how to set up the problem of image recognition, the learning algorithms (e.g. backpropagation), practical engineering tricks for training and fine-tuning the networks and guide the students through hands-on assignments and a final course project. Much of the background and materials of this course will be drawn from the ImageNet Challenge.

Course Instructors

Fei-Fei Li  
Andrej Karpathy

Teaching Assistants

Justin Johnson  
Yuke Zhu  
Brett Kuprel  
Ben Poole
These results are relatively impressive, but they are reasonable.
“The Unreasonable Effectiveness of Recurrent Neural Networks”

karpathy.github.io
Character-level language model example

Vocabulary: [h,e,l,o]

Example training sequence: “hello”

\[ h_{t+1} = \tanh(W_{hh}h_t + W_{xh}x_t) \]
Sonnet 116 – Let me not ...

by William Shakespeare

Let me not to the marriage of true minds
Admit impediments. Love is not love
Which alters when it alteration finds,
Or bends with the remover to remove:
O no! it is an ever-fixed mark
That looks on tempests and is never shaken;
It is the star to every wandering bark,
Whose worth's unknown, although his height be taken.
Love's not Time's fool, though rosy lips and cheeks
Within his bending sickle's compass come:
Love alters not with his brief hours and weeks,
But bears it out even to the edge of doom.
If this be error and upon me proved,
I never writ, nor no man ever loved.
PANDARUS:
Alas, I think he shall be come approached and the day
When little strain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:
Come, sir, I will make did behold your worship.

VIOLA:
Why, Salisbury must find his flesh and thought
That which I am not aps, not a man and in fire,
To show the reining of the raven and the wars
To grace my hand reproach within, and not a fair are hand,
That Caesar and my goodly father's world;
When I was heaven of presence and our fleets,
We spare with hours, but cut thy council I am great,
Murdered and by thy master's ready there
My power to give thee but so much as hell:
Some service in the noble bondman here,
Would show him to her wine.

KING LEAR:
O, if you were a feeble sight, the courtesy of your law,
Your sight and several breath, will wear the gods
With his heads, and my hands are wonder'd at the deeds,
So drop upon your lordship's head, and your opinion
Shall be against your honour.

VIOLA:
I'll drink it.
# The Stacks Project

## Browse chapters

<table>
<thead>
<tr>
<th>Part</th>
<th>Chapter</th>
<th>online</th>
<th>TeX source</th>
<th>view pdf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preliminaries</td>
<td>1. Introduction</td>
<td>online</td>
<td>tex</td>
<td>pdf</td>
</tr>
<tr>
<td></td>
<td>2. Conventions</td>
<td>online</td>
<td>tex</td>
<td>pdf</td>
</tr>
<tr>
<td></td>
<td>3. Set Theory</td>
<td>online</td>
<td>tex</td>
<td>pdf</td>
</tr>
<tr>
<td></td>
<td>4. Categories</td>
<td>online</td>
<td>tex</td>
<td>pdf</td>
</tr>
<tr>
<td></td>
<td>5. Topology</td>
<td>online</td>
<td>tex</td>
<td>pdf</td>
</tr>
<tr>
<td></td>
<td>6. Sheaves on Spaces</td>
<td>online</td>
<td>tex</td>
<td>pdf</td>
</tr>
<tr>
<td></td>
<td>7. Sites and Sheaves</td>
<td>online</td>
<td>tex</td>
<td>pdf</td>
</tr>
<tr>
<td></td>
<td>8. Stacks</td>
<td>online</td>
<td>tex</td>
<td>pdf</td>
</tr>
<tr>
<td></td>
<td>9. Fields</td>
<td>online</td>
<td>tex</td>
<td>pdf</td>
</tr>
<tr>
<td></td>
<td>10. Commutative Algebra</td>
<td>online</td>
<td>tex</td>
<td>pdf</td>
</tr>
</tbody>
</table>
Lemma 0.1. Assume (3) and (4) by the construction in the description.
Suppose \( X = \lim |X| \) by the formal open covering of \( X \) and a single map \( \text{Proj}_X(A) = \text{Spec}(B) \) over \( U \) compatible with the complex:

\[
\text{Set}(A) = \Gamma(X, \mathcal{O}_X, \mathcal{O}_X).
\]

When in this case of to show that \( S \rightarrow C_{Z/X} \) is stable under the following result in the second conditions of (1), and (3). This finishes the proof. By Definition ?? (without element is when the closed subschemes are catenary. If \( T \) is surjective we may assume that \( T \) is connected with residue fields of \( S \). Moreover there exists a closed subspace \( Z \subset X \) of \( X \) where \( U \) in \( X' \) is proper (some defining as a closed subset of the uniqueness it suffices to check that the following theorem

(1) \( f \) is locally of finite type. Since \( S = \text{Spec}(R) \) and \( Y = \text{Spec}(R) \).

Proof. This is form all sheaves of sheaves on \( X \). But given a scheme \( U \) and a surjective étale morphism \( U \rightarrow X \). Let \( U \cap U_i = \prod_{i=1}^n U_i \) be the scheme \( X \) over \( S \) at the schemes \( X_i \rightarrow X \) and \( U = \lim U_i \).

The following lemma surjective restrooms decomposes this implies that \( F_{x_0} = F_{x_0} \).

Lemma 0.2. Let \( X \) be a locally Noetherian scheme over \( S \); \( E = F_{X/S} \). Set \( I = J_1 < I_2 \). Since \( I \subset I_2 \) are nonzero over \( i_0 \leq p \) is a subset of \( J_0, \cdots, A_2 \) works.

Lemma 0.3. In Situation ??, Hence we may assume \( q' = 0 \).

Proof. We will use the property we see that \( p = \text{next functor} (??) \). On the other hand, by Lemma ?? we see that \( D(\mathcal{O}_X) = \mathcal{O}_X(D) \)

where \( K \) is an \( F \)-algebra where \( \delta_{n+1} \) is a scheme over \( S \).
Proof. Omitted.

Lemma 0.1. Let $C$ be a set of the construction.

Let $C$ be a gerber covering. Let $\mathcal{F}$ be a quasi-coherent sheaves of $\mathcal{O}$-modules. We have to show that

$$\mathcal{O}_{\mathcal{O}_X} = \mathcal{O}_X(\mathcal{L})$$

Proof. This is an algebraic space with the composition of sheaves $\mathcal{F}$ on $X_{\text{étale}}$ we have

$$\mathcal{O}_X(\mathcal{F}) = \{\text{morph}_1 \times \mathcal{O}_X(G, \mathcal{F})\}$$

where $G$ defines an isomorphism $\mathcal{F} \to \mathcal{F}$ of $\mathcal{O}$-modules.

Lemma 0.2. This is an integer $\mathcal{Z}$ is injective.

Proof. See Spaces, Lemma ??.

Lemma 0.3. Let $S$ be a scheme. Let $X$ be a scheme and $X$ is an affine open covering. Let $U \subset X$ be a canonical and locally of finite type. Let $X$ be a scheme.

Let $X$ be a scheme which is equal to the formal complex.

The following to the construction of the lemma follows.

Let $X$ be a scheme. Let $X$ be a scheme covering. Let

$$b : X \to Y' \to Y \to Y' \times_X Y \to X.$$ be a morphism of algebraic spaces over $S$ and $Y$.

Proof. Let $X$ be a nonzero scheme of $X$. Let $X$ be an algebraic space. Let $\mathcal{F}$ be a quasi-coherent sheaf of $\mathcal{O}_X$-modules. The following are equivalent

1. $\mathcal{F}$ is an algebraic space over $S$.
2. If $X$ is an affine open covering.

Consider a common structure on $X$ and $X$ the functor $\mathcal{O}_X(U)$ which is locally of finite type.

This since $\mathcal{F} \in \mathcal{F}$ and $x \in G$ the diagram

\[ S \xrightarrow{\xi} \mathcal{O}_X' \]

\[ \text{gor}_x \]

\[ \mathcal{O}_X' \]

\[ \alpha' \to \alpha \]

\[ X \]

\[ \text{Spec}(K_{\phi}) \]

\[ \text{Mor}_{\text{Sets}}(d(\mathcal{O}_X, G)) \]

is a limit. Then $G$ is a finite type and assume $S$ is a flat and $\mathcal{F}$ and $G$ is a finite type $\mathcal{F}_0$, This is of finite type diagrams, and

- the composition of $G$ is a regular sequence,
- $\mathcal{O}_X$ is a sheaf of rings.

Proof. We have see that $X = \text{Spec}(R)$ and $\mathcal{F}$ is a finite type representable by algebraic space. The property $\mathcal{F}$ is a finite morphism of algebraic stacks. Then the cohomology of $X$ is an open neighbourhood of $U$.

Proof. This is clear that $G$ is a finite presentation, see Lemmas ??.

A reduced above we conclude that $U$ is an open covering of $C$. The functor $\mathcal{F}$ is a field

$$\mathcal{O}_{X,x} \to \mathcal{F}_x \to \mathcal{O}_X(\mathcal{O}_{X,x})$$

is an isomorphism of covering of $\mathcal{O}_{X,x}$. If $\mathcal{F}$ is the unique element of $\mathcal{F}$ such that $X$ is an isomorphism.

The property $\mathcal{F}$ is a disjoint union of Proposition ?? and we can filtered set of presentations of a scheme $\mathcal{O}_X$-algebra with $\mathcal{F}$ are open of finite type over $S$.

If $\mathcal{F}$ is a scheme theoretic image points.

If $\mathcal{F}$ is a finite direct sum $\mathcal{O}_{X,x}$ is a closed immersion, see Lemma ??, This is a sequence of $\mathcal{F}$ is a similar morphism.
## Linux Kernel Source Tree

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documentation</td>
<td>Merge git://git.kernel.org/pub/scm/linux/kernel/git/nab/target-pending</td>
<td>6 days ago</td>
</tr>
<tr>
<td>arch</td>
<td>Merge branch 'x86-urgent-for-linus' of git://git.kernel.org/pub/scm/l...</td>
<td>a day ago</td>
</tr>
<tr>
<td>block</td>
<td>block: discard bdi_unregister() in favour of bdi_destroy()</td>
<td>9 days ago</td>
</tr>
<tr>
<td>crypto</td>
<td>Merge git://git.kernel.org/pub/scm/linux/kernel/git/herbert/crypto-2.6</td>
<td>10 days ago</td>
</tr>
<tr>
<td>drivers</td>
<td>Merge branch 'drm-fixes' of git://people.freedesktop.org/~airlied/linux</td>
<td>9 hours ago</td>
</tr>
<tr>
<td>firmware</td>
<td>firmware/ihex2fw.c: restore missing default in switch statement</td>
<td>2 months ago</td>
</tr>
<tr>
<td>fs</td>
<td>vfs: read file_handle only once in handle_to_path</td>
<td>4 days ago</td>
</tr>
<tr>
<td>include</td>
<td>Merge branch 'per-urgent-for-linus' of git://git.kernel.org/pub/scm/...</td>
<td>a day ago</td>
</tr>
<tr>
<td>init</td>
<td>init: fix regression by supporting devices with major:minor:offset...</td>
<td>a month ago</td>
</tr>
<tr>
<td>io</td>
<td>Merge branch 'fsc-linux' of git://git.kernel.org/pub/scm/fsc/Linux</td>
<td>a month ago</td>
</tr>
</tbody>
</table>
static void do_command(struct seq_file *m, void *v)
{
    int column = 32 << (cmd[2] & 0x80);
    if (state)
        cmd = (int)(int_state ^ (in_8(&ch->ch_flags) & Cmd) ? 2 : 1);
    else
        seq = 1;
    for (i = 0; i < 16; i++) {
        if (k & (1 << i))
            pipe = (in_use & UMXTTHREAD_UNCCA) +
                  ((count & 0x00000000fffffff8) & 0x000000f) << 8;
        if (count == 0)
            sub(pid, ppc_md.kexec_handle, 0x20000000);
        pipe_set_bytes(i, 0);
    }
    /* Free our user pages pointer to place camera if all dash */
    subsystem_info = &of_changes[PAGE_SIZE];
    rek_controls(offset, idx, &soffset);
    /* Now we want to deliberately put it to device */
    control_check_polarity(&context, val, 0);
    for (i = 0; i < COUNTER; i++)
        seq_puts(s, "policy ");
}
static void do_command(struct seq_file *m, void *v) {
    int column = 32 << (cmd[2] & 0x80);
    if (state)
        cmd = (int)(int_state ^ (in_8(&ch->ch_flags) & Cmd) ? 2 : 1);
    else
        seq = 1;
    for (i = 0; i < 16; i++) {
        if (k & (1 << i))
            pipe = (in_use & UMPTHREAD_UNCCA) +
            ((count & 0x00000000fffffff8) & 0x000000f) << 8;
        if (count == 0)
            sub(pid, ppc_md.kexec_handle, 0x20000000);
        pipe_set_bytes(i, 0);
    }
    /* Free our user pages pointer to place camera if all dash */
    subsystem_info = &of_changes[PAGE_SIZE];
    rek_controls(offset, idx, &osoffset);
    /* Now we want to deliberately put it to device */
    control_check_polarity(&context, val, 0);
    for (i = 0; i < COUNTER; i++)
        seq_puts(s, "policy ");
}
/*
 * Copyright (c) 2006-2010, Intel Mobile Communications. All rights reserved.
 *
 * This program is free software; you can redistribute it and/or modify it
 * under the terms of the GNU General Public License version 2 as published by
 * the Free Software Foundation.
 *
 * This program is distributed in the hope that it will be useful,
 * but WITHOUT ANY WARRANTY; without even the implied warranty of
 * MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
 *
 * GNU General Public License for more details.
 *
 * You should have received a copy of the GNU General Public License
 * along with this program; if not, write to the Free Software Foundation,
 * Inc., 675 Mass Ave, Cambridge, MA 02139, USA.
 */

#include <linux/kexec.h>
#include <linux/errno.h>
#include <linux/io.h>
#include <linux/platform_device.h>
#include <linux/multi.h>
#include <linux/ckevent.h>

#include <asm/io.h>
#include <asm/prom.h>
#include <asm/e820.h>
#include <asm/system_info.h>
#include <asm/setew.h>
#include <asm/pgproto.h>
#include <asm/io.h>
#include <asm/prom.h>
#include <asm/e820.h>
#include <asm/system_info.h>
#include <asm/setev.h>
#include <asm/pgproto.h>

#define REG_PG vesa_slot_addr_pack
#define PFM_NOCOMP AFSR(0, load)
#define STACK_DDR(type) ((func)

#define SWAP_ALLOCATE(nr) (e)
#define emulate_sigs() arch_get_unaligned_child()
#define access_rw(TST) asm volatile("movd %$esp, %0, %3", "r"(0));

static void stat_PC_SEC __read_mostly offsetof(struct seq_argsqueue, pC[1]);

static void
os_prefix(unsigned long sys)
{
#ifdef CONFIG_PREEMPT
    PUT_PARAM_RAID(2, sel) = get_state_state();
    set_pid_sum((unsigned long)state, current_state_str(), (unsigned long)-1->lr_full, low;
}
Try it yourself: **char-rnn** on Github (uses Torch7)
Title: BASIC CHEESE WINGS:
Categories: Desserts
Yield: 6 Servings

3 Eggs
2 tb Chopped fresh curry
-or cooking spray
1 c Water; cooked
2 Lemons minced mushrooms
3 oz Sweet cooked rice
1/2 Onion; chopped
3 c Butter, melted
2 ts Soy sauce
1 ts Cinnamon
2 md Sugar or food coloring;
-stems cored bowl
2 tb Salt and freshly grated
1/4 ts Ground ginger
1/2 c Flour
1 tb Water; fresh parsley
1 c Water (or or)
1 Clove garlic, minced

Preheat oven to 350F. Combine sugar, salt, baking soda, celery and sugar. Add the chicken broth well. Add the cornstarch to the pan; cool. Add the olive oil, oil, and basil or cooking spray. Pour the onions until melted.
Good afternoon. God bless you.

The United States will step up to the cost of a new challenges of the American people that will share the fact that we created the problem. They were attacked and so that they have to say that all the task of the final days of war that I will not be able to get this done. The promise of the men and women who were still going to take out the fact that the American people have fought to make sure that they have to be able to protect our part. It was a chance to stand together to completely look for the commitment to borrow from the American people. And the fact is the men and women in uniform and the millions of our country with the law system that we should be a strong strectks of the forces that we can afford to increase our spirit of the American people and the leadership of our country who are on the Internet of American lives.

Thank you very much. God bless you, and God bless the United States of America.
1:15 And the LORD said unto Moses, Thus shall it come to pass, that a gate in the land of the nations, to slay the third day.

2:34 And the LORD shall smite them in the day of judgment, and to bars, and it shall devour them; and I will consume them.

22:14 And he said, They shall be given unto them that shouted me: and there was a god of innocent blue, in the days of the family of Jacob.
Music,

ABC notation:

T:Lat canny I. the dlas.
M:C
L:1/8
Q:1/2=100
K:D
A\>|F>DEA|F2dF|A/F/A/B/ A/F/G/E/E FD |DDD|G EdEd|efd| leged|f|g|d cd2||
|el|g2|ef| ge(f|e|c/)| l/dd|fe fdAA|F3 A c4|e|f|e|f|g{e}|d4 |
gfga afgfle|ggbb ad'eg|f|gdB edAB|B|edA BABg|f|d|d|e ddd:|
### NSF Abstracts

**Title**: Mathematical Sciences: An Integration Diffusivity in Mechanism of Processing and Minimal Components in Central Topology

**Type**: Award

**NSF Org**: DUE

**Date**: April 11, 1996

**File**: a9455932

**Award Number**: 9455924

**Award Instr.**: Standard Grant

**Prgm Manager**: Stephan P. Nelson

**DUE DIVISION OF UNDERGRADUATE EDUCATION**

**EHR DIRECTORATE FOR ENGINEERING**

**Start Date**: March 1, 1999

**Expires**: February 28, 2001 (Estimated)

**Expected Total Amt.**: $150000 (Estimated)

**Principal Investigator**: Jennifer E. Strislon

**Sponsor**: U of Cal Davis

**OVCR/Sponsor/Program Survey**

**Chicago, IL 60637-1404 788/624-3111**

**NSF Program**: 1155 BIOMOLECULAR PROCASSN

**Fld Applcns**: 0000099 Other Applications NEC

**Program Ref**: 0000, OTHER

**Abstract**:

Decreasing a single international representation of the forces of protein collapse and preservations, and mathematically important next links of basis by the programming central the development of the circuit traces in all the productional materials. Recent years are highly significant from professional elements. The proposed research will examine the cortical levels of the effect of sedimentation. The properties of the theory of these hyperpolarization criterions will be examined. In particular, we are assessed by the availability of natural materials.
Anaboth Cubblue

Artifact

\[\text{\textbf{2}, \text{\textcolor{white}{\textbf{1}}}}: \text{Target land becomes a 4/4 white Spirit creature with flying until end of turn.}\]

Roon War Medoma

\[\text{\textbf{2}, \text{\textcolor{white}{\textbf{1}}}}: \text{Name a card. You gain 1 life.}\]

Blood Rhast

\[\text{\textbf{1}, \text{\textcolor{white}{\textbf{1}}}}: \text{Blood Rhast gains "Sacrifice a land: Blood Rhast deals 3 damage to target player."}\]

\[2/2\]

Rakile Volan

\[\text{\textbf{1}, \text{\textcolor{white}{\textbf{1}}}}: \text{Whenever Rakile Volan blocks or becomes blocked, you may pay 1. If you do, you gain 2 life.}\]

\[1/1\]
**Roon War Medoma**

*Instant*

Name a card. You gain 1 life.

---

**Light of the Bild**

*Creature – Spirit*

Flying

Whenever Light of the Bild blocks, you may put a 1/1 green Angel creature token with flying onto the battlefield.

*Creature – Dragon*

Mountainspalk

Trompel (If this creature would assign enough damage to its blockers to destroy them, it assigns the rest of its damage to you.)

---

**Slidshocking Krow**

*Creature – Spirit*

Whenever Rakile Volan blocks or becomes blocked, you may pay 1 {C}. If you do, you gain 2 life.

*Creature – Dragon*
APEMANTUS:
Ho, ho! I laugh to see your beard!

BOYET:
Madam, in great extremes of passion as she discovers it.

PAROLLES:
By my white head and her wit
Values itself: to the sepulchre!
With this, my lord,
That I have some business: let's away.

First Keeper:
Forbear to murder: and wilt thou not say he lies,
And lies, and let the devil would have said, sir, their speed Hath been balm to heal their woes,
But on Linux:

```c
~/*
 * linux/kernel/time.c
 * Please report this on hardware.
 */

void irq_mark_irq(unsigned long old_entries, eval);

/*
   * Divide only 1000 for ns^2 -> us^2 conversion values
don't overflow:
   seq_puts(m, \"\ttramp: %ps\",
   (void *)class->contending_point
]++;
   if (likely(t->flags & WQ_UNBOUND)) {
      /*
         * Update inode information. If the
         * slowpath and sleep time (abs or rel)
         * @rmp: remaining (either due
         * to consume the state of ring buffer size. */
         header_size - size, in bytes, of the chain.
      */
      BUG_ON(!error);
   } while (cgrp) {
      if (old) {
         if (kdb_continue_catastrophic;
##endif
```
Visualizing and Understanding Recurrent Networks
Andrej Karpathy*, Justin Johnson*, Li Fei-Fei
(Arxiv as of 3 weeks ago)
RNN:

\[ h_t^l = \tanh W^l \left( \begin{pmatrix} h_t^{l-1} \\ h_t^{l-1} \end{pmatrix} \right) \]

\[ h \in \mathbb{R}^n, \quad W^l [n \times 2n] \]
LSTM

LSTM: A Search Space Odyssey, Greff et al.
RNN:

\[ h^l_t = \tanh W^l \left( h^l_{t-1} \right) \]

\[ h \in \mathbb{R}^n \quad W^l \left[ n \times 2n \right] \]

LSTM:

\[ \begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \text{sigm} \\ \text{sigm} \\ \text{sigm} \\ \text{tanh} \end{pmatrix} W^l \left( h^l_{t-1} \right) \]

\[ c^l_t = f \odot c^l_{t-1} + i \odot g \]

\[ h^l_t = o \odot \tanh(c^l_t) \]

\[ W^l \left[ 4n \times 2n \right] \]

depth

time
Datasets

Leo Tolstoy’s “War and Peace”

This black-eyed, wide-mouthed girl, not pretty but full of life—with childish bare shoulders which after her run heaved and shook her bodice, with black curls tossed backward, thin bare arms, little legs in lace-frilled drawers, and foot in low slippers—was just at that charming age when a girl is no longer a child, though the child is not yet a young woman. Escaping from her father she ran to hide her flushed face in the lace of her mother’s mantilla—not paying the least attention to her severe remark—and began to laugh. She laughed, and in fragmentary sentences tried to explain about a doll which she produced from the folds of her frock.

"Do you see?... My doll... Mimi... You see..." was all Natasha managed to utter (to her everything seemed funny). She leaned against her mother and burst into such a loud, ringing fit of laughter that even the prim visitor could not help joining in.

"Now then, go away and take your monstrosity with you," said the mother, pushing away her daughter with pretended sternness, and turning to the visitor she added: "She is my youngest girl."

Natasha, raising her face for a moment from her mother’s mantilla, glanced up at her through tears of laughter, and again hid her face.

The visitor, compelled to look on at this family scene, thought it necessary to take some part in it.

"Tell me, my dear," said she to Natasha, "is Mimi a relation of yours? A daughter, I suppose?"

Natasha did not like the visitor’s tone of condescension to childish things. She did not reply, but looked at her seriously.

Linus Torvald’s “Linux Kernel”

```
static ssize_t
sched_fstat_write(struct file *filp, const char __user *ubuf,
                   size_t cnt, loff_t *ppos)
{
    char buf[64];
    char *cmp;
    int i;
    struct inode *inode;

    if (cnt > 63)
        cnt = 63;

    if (copy_from_user(&buf, ubuf, cnt))
        return -EFAULT;

    buf[cnt] = 0;
    cmp = strlen(buf);

    /* Ensure the static key remains in a consistent state */
    inode = file_inode(filp);
    mutex_lock(&inode->__i_mutex);
    i = sched_fstat_set(cmp);
    mutex_unlock(&inode->__i_mutex);
    if (i == SCHED_FSTAT_NN)
        return -ENOSYS;

    *ppos += cnt;
    return cnt;
}
```
### Step 1: Performance

#### War and Peace Dataset

<table>
<thead>
<tr>
<th>Model</th>
<th>n</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-gram</td>
<td></td>
<td>2.399</td>
<td>1.928</td>
<td>1.521</td>
<td>1.314</td>
<td>1.232</td>
<td>1.203</td>
<td><strong>1.194</strong></td>
<td>1.194</td>
<td>1.194</td>
<td>1.195</td>
</tr>
<tr>
<td>n-NN</td>
<td></td>
<td>2.399</td>
<td>1.931</td>
<td>1.553</td>
<td>1.451</td>
<td>1.339</td>
<td><strong>1.321</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

#### Linux Kernel Dataset

<table>
<thead>
<tr>
<th>Model</th>
<th>n</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-gram</td>
<td></td>
<td>2.702</td>
<td>1.954</td>
<td>1.440</td>
<td>1.213</td>
<td>1.097</td>
<td>1.027</td>
<td>0.982</td>
<td>0.953</td>
<td>0.933</td>
<td><strong>0.889</strong></td>
</tr>
<tr>
<td>n-NN</td>
<td></td>
<td>2.707</td>
<td>1.974</td>
<td>1.505</td>
<td>1.395</td>
<td><strong>1.256</strong></td>
<td>1.376</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
### Step 1: Performance

<table>
<thead>
<tr>
<th>Model</th>
<th>( n )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>War and Peace Dataset</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n )-gram</td>
<td>2.399</td>
<td>1.928</td>
<td>1.521</td>
<td>1.314</td>
<td>1.232</td>
<td>1.203</td>
<td><strong>1.194</strong></td>
<td>1.194</td>
<td>1.194</td>
<td>1.195</td>
<td></td>
</tr>
<tr>
<td>( n )-NN</td>
<td>2.399</td>
<td>1.931</td>
<td>1.553</td>
<td>1.451</td>
<td>1.339</td>
<td><strong>1.321</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td><strong>Linux Kernel Dataset</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n )-gram</td>
<td>2.702</td>
<td>1.954</td>
<td>1.440</td>
<td>1.213</td>
<td>1.097</td>
<td>1.027</td>
<td>0.982</td>
<td>0.953</td>
<td>0.933</td>
<td><strong>0.889</strong></td>
<td></td>
</tr>
<tr>
<td>( n )-NN</td>
<td>2.707</td>
<td>1.974</td>
<td>1.505</td>
<td>1.395</td>
<td><strong>1.256</strong></td>
<td>1.376</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Layers</th>
<th>LSTM</th>
<th>RNN</th>
<th>GRU</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>War and Peace Dataset</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>1.449</td>
<td>1.442</td>
<td>1.540</td>
</tr>
<tr>
<td>128</td>
<td>1.277</td>
<td><strong>1.227</strong></td>
<td>1.279</td>
</tr>
<tr>
<td>256</td>
<td>1.189</td>
<td><strong>1.137</strong></td>
<td>1.141</td>
</tr>
<tr>
<td>512</td>
<td>1.161</td>
<td>1.092</td>
<td>1.082</td>
</tr>
<tr>
<td><strong>Linux Kernel Dataset</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>1.355</td>
<td><strong>1.331</strong></td>
<td>1.366</td>
</tr>
<tr>
<td>128</td>
<td>1.149</td>
<td><strong>1.128</strong></td>
<td>1.177</td>
</tr>
<tr>
<td>256</td>
<td>1.026</td>
<td><strong>0.972</strong></td>
<td>0.998</td>
</tr>
<tr>
<td>512</td>
<td>0.952</td>
<td>0.840</td>
<td>0.846</td>
</tr>
</tbody>
</table>
## Step 1: Performance

<table>
<thead>
<tr>
<th>Model</th>
<th>( n )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n )-gram</td>
<td>1</td>
<td>2.399</td>
<td>1.928</td>
<td>1.521</td>
<td>1.314</td>
<td>1.232</td>
<td>1.203</td>
<td>1.194</td>
<td>1.194</td>
<td>1.194</td>
<td>1.195</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.399</td>
<td>1.931</td>
<td>1.553</td>
<td>1.451</td>
<td>1.339</td>
<td>1.321</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n )-NN</td>
<td>1</td>
<td>2.702</td>
<td>1.954</td>
<td>1.440</td>
<td>1.213</td>
<td>1.097</td>
<td>1.027</td>
<td>0.982</td>
<td>0.953</td>
<td>0.933</td>
<td>0.889</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.707</td>
<td>1.974</td>
<td>1.505</td>
<td>1.395</td>
<td>1.256</td>
<td>1.376</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### War and Peace Dataset

### Linux Kernel Dataset

<table>
<thead>
<tr>
<th>Layers</th>
<th>LSTM</th>
<th>RNN</th>
<th>GRU</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Size</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

### Size

<table>
<thead>
<tr>
<th>War and Peace Dataset</th>
<th>Linux Kernel Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.449</td>
<td>1.189</td>
</tr>
<tr>
<td>1.442</td>
<td>1.137</td>
</tr>
<tr>
<td>1.540</td>
<td>1.141</td>
</tr>
<tr>
<td>1.446</td>
<td>1.342</td>
</tr>
<tr>
<td>1.401</td>
<td>1.256</td>
</tr>
<tr>
<td>1.396</td>
<td>1.170</td>
</tr>
<tr>
<td>1.472</td>
<td>1.201</td>
</tr>
</tbody>
</table>

### 3GB Models

### 11MB Models
Linux
Linux

War and Peace
Closing brace ("\{\}) case study
Searching for interpretable cells

\[
\begin{pmatrix}
i \\
f \\
o \\
g
\end{pmatrix} =
\begin{pmatrix}
sigm \\
sigm \\
tanh
\end{pmatrix}
W^l
\begin{pmatrix}
h_{t-1}^{l-1} \\
h_t^l \\
h_t^{l-1}
\end{pmatrix}
\]

\[
c_t^l = f \odot c_{t-1}^l + i \odot g
\]

\[
h_t^l = o \odot \tanh(c_t^l)
\]
Searching for interpretable cells

/* Unpack a filter field's string representation from user-space buffer. */
char* audit_unpack_string(void **buftp, size_t *remain, size_t len)
{
    char *str;
    if (!*buftp || (len == 0) || (len > *remain))
        return ERR_PTR(-EINVAL);

    /* Of the currently implemented string fields, PATH_MAX defines the longest valid length. */
"You mean to imply that I have nothing to eat out of.... On the contrary, I can supply you with everything even if you want to give dinner parties," warmly replied Chichagov, who tried by every word he spoke to prove his own rectitude and therefore imagined Kutuzov to be animated by the same desire.

Kutuzov, shrugging his shoulders, replied with his subtle penetrating smile: "I meant merely to say what I said."
Searching for interpretable cells

"You mean to imply that I have nothing to eat out of.... On the contrary, I can supply you with everything even if you want to give dinner parties," warmly replied Chichagov, who tried by every word he spoke to prove his own rectitude and therefore imagined Kutuzov to be animated by the same desire.

Kutuzov, shrugging his shoulders, replied with his subtle penetrating smile: "I meant merely to say what I said."

Interesting note: backpropagation horizon was only 100 time steps
Searching for interpretable cells

Cell sensitive to position in line:
The sole importance of the crossing of the Berezina lies in the fact that it plainly and indubitably proved the fallacy of all the plans for cutting off the enemy's retreat and the soundness of the only possible line of action—the one Kutuzov and the general mass of the army demanded—namely, simply to follow the enemy up. The French crowd fled at a continually increasing speed and all its energy was directed to reaching its goal. It fled like a wounded animal and it was impossible to block its path. This was shown not so much by the arrangements it made for crossing as by what took place at the bridges. When the bridges broke down, unarmed soldiers, people from Moscow and women with children who were with the French transport, all—carried on by vis inertiae—pressed forward into boats and into the ice-covered water and did not, surrender.

line length tracking cell
Searching for interpretable cells

The sole importance of the crossing of the Berezina lies in the fact that it plainly and indubitably proved the fallacy of all the plans for cutting off the enemy's retreat and the soundness of the only possible line of action—the one Kutuzov and the general mass of the army demanded—namely, simply to follow the enemy up. The French crowd fled at a continually increasing speed and all its energy was directed to reaching its goal. It fled like a wounded animal and it was impossible to block its path. This was shown not so much by the arrangements it made for crossing as by what took place at the bridges. When the bridges broke down, unarmed soldiers, people from Moscow and women with children who were with the French transport, all—carried on by vis inertiae—pressed forward into boats and into the ice-covered water and did not, surrender.
Searching for interpretable cells

```c
static int __dequeue_signal(struct sigpending *pending, sigset_t *mask, siginfo_t *info)
{
    int sig = next_signal(pending, mask);
    if (sig) {
        if (current->notifier) {
            if (sigismember(current->notifier_mask, sig)) {
                if (current->notifier)(current->notifier_data)
                    clear_thread_flag(TIF_SIGPENDING);
                return 0;
            }
        }
    }
    collect_signal(sig, pending, info);
    return sig;
}
```

if statement cell
Searching for interpretable cells

```c
/* Duplicate LSM field information. The lsm_rule is opaque, so
 * re-initialized. */
static inline int audit_dupe_lsm_field(struct audit_field *df,
    struct audit_field *sf)
{
    int ret = 0;
    char *lsm_str;
    /* our own copy of lsm_str */
    lsm_str = kstrdup(sf->lsm_str, GFP_KERNEL);
    if (unlikely(!lsm_str))
        return -ENOMEM;
    df->lsm_str = lsm_str;
    /* our own (refreshed) copy of lsm_rule */
    ret = securityaudit_rule_init(df->type, df->op, df->lsm_str,
        (void **) &df->lsm_rule);
    /* Keep currently invalid fields around in case they
     * become valid after a policy reload. */
    if (ret == -EINVAL) {
        pr_warn("audit rule for LSM '%s' is invalid\n",
            df->lsm_str);
        ret = 0;
    }
    return ret;
}
```
Searching for interpretable cells

```c
#define CONFIG_AUDITSYSCALL

static inline int audit_match_class_bits(int class, u32 *mask)
{
    int i;
    if (classes[class]) {
        for (i = 0; i < AUDIT_BITMASK_SIZE; i++)
            if (mask[i] & classes[class][i])
                return 0;
    }
    return 1;
}
```
Searching for interpretable cells

```c
char *audit_unpack_string(void **bufp, size_t *remain, si
{
    char *str;
    if (!*bufp || (len == 0) || (len > *remain))
        return ERR_PTR(-EINVAL);
    /* of the currently implemented string fields, PATH_MAX
    * defines the longest valid length.
    */
    if (len > PATH_MAX)
        return ERR_PTR(-ENAMETOOLONG);
    str = kmalloc(len + 1, GFP_KERNEL);
    if (unlikely(!str))
        return ERR_PTR(-ENOMEM);
    memcpy(str, *bufp, len);
    str[len] = 0;
    *bufp += len;
    *remain -= len;
    return str;
}
```
Learning Dynamics

Training Iterations

"Tmont thithey" fomescerliund
Keushey. Thom here
sheulke, ammerenith ol sivh I lalterthend Bleipile shuwy fil on aselterlome
coaniogennc Phe lism thond hon at. MeIDimoration in ther thize."

"Kite vouch!" he repeated by her
door. "But I would be done and quarts, feeling, then, son is people...."

"Why do what that day," replied Natasha, and wishing to himself the fact the
princess, Princess Mary was easier, fed in had oftened him.
Pierre aking his soul came to the packs and drove up his father-in-law women.
Learning Dynamics

Training iterations

"Tmont thithey" fomesscerliund
Keushey. Thom here
shulke, anmerenith ol sivh I latlerthend Bleipile shuwy fil on aseterlome
cooniogenncc Phe lism thond hon at. MeiDimortion in ther thize."

"We counter. He stutn co des. His stanted out one ofler that co
to gearang reay Jotrets and with fre colt otf paitt thin wall.

"Aftair fall unsuch that the hall for Phe
her hearly, and behs to so arwage fivin
how, and Gognition is so overetical and

"Kite vouch!" he repeated by her
door. "But I would be done and quarts, feeling, then, son is

"Why do what that day," replied Natasha
princess, Princess Mary was easier, feet
Pierre aking his soul came to the packs

1-NN  3-NN
2-NN  4-NN

mean KL

LSTM epoch
Learning Dynamics

Sequence to Sequence Learning with Neural Networks,
Sutskever et al.
Error Analysis
n-gram oracle

Optimistic estimate of how many errors could be eliminated by better modeling the last n characters:

- remove error if correctly classified by ANY ngram model, for n = 1 .. 9

1-gram to 5-gram

"My wife," continued Prince Andrew, "is an excellent woman, one of those rare women with whom a man's honor is safe; but, O God, what would I not give now to be unmarried! You are the first and only one to whom I mention this, because I like you."

18% of errors
dynamic memory oracle

“Jon yelled at Mary but Mary couldn’t hear him”
dynamic memory oracle

Remove errors for words that just occurred within the last n characters. (n = 100, 500, 1000, 5000)

Up to 500 memory

circular, memorandum, or report, skillfully, pointedly, and elegantly. Bilibin's services were valued not only for what he wrote, but also for his skill in dealing and conversing with those in the highest spheres.

Bilibin liked conversation as he liked work, only when it could be made elegantly witty. In society he always awaited an opportunity to say something striking and took part in a conversation only when that was possible. His conversation was always sprinkled with wittily original,

6% of errors
rare words oracle

Remove errors for words that occur very infrequently in the training data (n = 0...5).

Less than 3 training examples of word

Nicholas and Sonya, the niece. Sonya was a slender little brunette with a tender look in her eyes which were veiled by long lashes, thick black plaits coiling twice round her head, and a tawny tint in her complexion and especially in the color of her slender but graceful and muscular arms and neck. By the grace of her movements, by the softness and flexibility of her small limbs, and by a certain coyness and reserve of
difficult next letter oracles

After space, quote, new line

"No, impossible!" said Prince Andrew, laughing and pressing Pierre's hand to show that there was no need to ask the question. He wished to

Anna Pavlovna smiled and promised to take Pierre in hand. She knew his father to be a connection of Prince Vasili's. The elderly lady who had been sitting with the old aunt rose hurriedly and overtook Prince Vasili

37% of errors
Smaller model makes 44K more errors: 36K of these are n-gram errors! (81%) 5K boost 3K distributed among rest

=> Scaling the model up gets rid of mainly n-gram errors and leaves the other error types almost untouched by comparison!
Conclusions

- LSTMs are powerful models and do learn interesting, interpretable, long-term interactions

- Limitations:
  - n-gram failures: fixable with scaling up the model
  - rare word failures: scale up data / transfer learning
  - dynamic memory errors: ??? (memory nets?)
  - word-level errors: hierarchies? clockwork RNN? not clear
Thank you!