시각 Tech. 심화과정 (3)

연세대학교 컴퓨터과학과

김 선 주

SK New ICT Academy
- AI Course -

AGENDA

- Video Processing
- Geometry (Optical Flow, Stereo)

Recurrent Neural Network

Slides from Stanford CS231n Fei Fei Li et al.

Recurrent Neural Network

- RNN's are very useful for NLP, in modelling the language model.
- The applications of language models are two-fold:
  - It allows us to score arbitrary sentences based on how likely they are to occur in the real world.
    This gives us a measure of grammatical and semantic correctness. Such models are typically used as part of Machine Translation systems.
  - A language model allows us to generate new text.
    Training a language model on Shakespeare allows us to generate Shakespeare-like text.

Recurrent Neural Network

- The idea behind RNNs is to make use of sequential information. In a traditional neural network we assume that all inputs (and outputs) are independent of each other.
  But for many tasks that’s a very bad idea. If you want to predict the next word in a sentence you better know which words came before it.

Recurrent Neural Network

- RNNs are called *recurrent* because they perform the same task for every element of a sequence, with the output being depended on the previous computations.
- Another way to think about RNNs is that they have a “memory” which captures information about what has been calculated so far.
- In theory RNNs can make use of information in arbitrarily long sequences, but in practice they are limited to looking back only a few steps (more on this later).

Recurrent Neural Network

A recurrent neural network and the unfolding in time of the computation involved in its forward computation. Source: Nature

Recurrent Neural Network

- x_t is the input at time step t. For example, x_1 could be a one-hot vector corresponding to the second word of a sentence.
- s_t is the hidden state at time step t. It's the "memory" of the network. s_t is calculated based on the previous hidden state and the input at the current step: $s_t = f(Ux_t + Ws_{t-1})$. The function f usually is a nonlinearity such as tanh or ReLU. s_{-1}, which is required to calculate the first hidden state, is typically initialized to all zeroes.
- o_t is the output at step t. For example, if we wanted to predict the next word in a sentence it would be a vector of probabilities across our vocabulary. $o_t = \mathrm{softmax}(Vs_t)$.

Recurrent Neural Network

- You can think of the hidden state s_t as the memory of the network. s_t captures information about what happened in all the previous time steps. The output at step o_t is calculated solely based on the memory at time t. As briefly mentioned above, it's a bit more complicated in practice because s_t typically can't capture information from too many time steps ago.
- Unlike a traditional deep neural network, which uses different parameters at each layer, a RNN shares the same parameters (U, V, W above) across all steps. This reflects the fact that we are performing the same task at each step, just with different inputs. This greatly reduces the total number of parameters we need to learn.
- The above diagram has outputs at each time step, but depending on the task this may not be necessary. For example, when predicting the sentiment of a sentence we may only care about the final output, not the sentiment after each word. Similarly, we may not need inputs at each time step. The main feature of an RNN is its hidden state, which captures some information about a sequence.

Recurrent Networks offer a lot of flexibility:

Vanilla Neural Networks

Recurrent Networks offer a lot of flexibility:

e.g. **Image Captioning**
image -> sequence of words

Recurrent Networks offer a lot of flexibility:

e.g. **Sentiment Classification**
sequence of words -> sentiment

Recurrent Networks offer a lot of flexibility:

e.g. **Machine Translation**
seq of words -> seq of words

http://www.cs.toronto.edu/~ilya/rnn.html

Recurrent Networks offer a lot of flexibility:

e.g. **Video classification on frame level**

RNN Extensions

- **Bidirectional RNNs** are based on the idea that the output at time may not only depend on the previous elements in the sequence, but also future elements. For example, to predict a missing word in a sequence you want to look at both the left and the right context. Bidirectional RNNs are quite simple. They are just two RNNs stacked on top of each other. The output is then computed based on the hidden state of both RNNs.

RNN Extensions

- **Deep (Bidirectional) RNNs** are similar to Bidirectional RNNs, only that we now have multiple layers per time step. In practice this gives us a higher learning capacity (but we also need a lot of training data)

RNN Extensions

- **LSTM networks** are quite popular these days and we briefly talked about them above. LSTMs don't have a fundamentally different architecture from RNNs, but they use a different function to compute the hidden state. The memory in LSTMs are called *cells* and you can think of them as black boxes that take as input the previous state and current input . Internally these cells decide what to keep in (and what to erase from) memory. They then combine the previous state, the current memory, and the input. It turns out that these types of units are very efficient at capturing long-term dependencies.

What can Language Model do?

- Translation
- Generation

http://www.cs.toronto.edu/~ilya/rnn.html

https://magenta.tensorflow.org/sketch-rnn-demo

<p class="cleartext"> Products: Laser Printers The fundamental everyday requirement for mono and colour laser printing throughtout today's offices is perfectly met with the extensive Epson laser printer range. The latest AcuLaser printer range offers users exceptionally Epson AcuLaser C1900 Networked compact colour laser printer for professional enterprises Businesses have been denied simple and affordable colour laser printing for far too long. The traditionally high costs and poor speeds of colour lasers has left many offices looking a bit, well, grey. But not any more: with the Epson AcuLaser C1900, Epson brings both colour and monochrome laser printing together at a black and white price. more Where to Buy Support Epson AcuLaser C3000 The fastest colour laser printer in its class The perfect printer for small businesses and work groups, the Epson AcuLaser C3000 prints high volumes in black and white and vibrant colour, at high speed and with low running costs.. more Where to Buy High quality resolution: 2400dpi RIT* Large paper capacity: 600 sheets, expandable up to 1,600 sheets Compatible Windows and Mac High speed USB and EpsonNet 10/100 Base Tx Ethernet interfaces as standard** * Epson AcuLaser Resolution Improvement Technology **EpsonNet 10/100 Base Tx Ethernet standard with Epson AcuLaser C3000N model only: AcuLaser C3000: 64MB Memory, 100 sheet MP Tray, 500 sheet cassette, Duplex printing as standard AcuLaser C3000N: 64MB Memory, 100 sheet MP Tray, 500 sheet cassette, Duplex printing, 10/100BaseTX Ethernet Interface Networked compact colour laser printer for professional enterprises Businesses have been denied simple and affordable colour laser printing for far too long. The traditionally high costs and poor speeds of colour lasers has left many offices looking a bit, well, grey. But not any more; with the Epson AcuLaser C1900, Epson brings both colour and monochrome laser printing together at a black and white price. Key Features cost effective mono printing for day to day business needs and vivid versatile colour when required. search Search Epson UK Epson AcuLaser C900 Outstanding professional colour printing for business Add colour to your business with the Epson AcuLaser C900 from Epson. Its perfect for the smaller workgroup, being a compact and cost effective laser printing workhorse that offers amazing colour output as well as high performance black and white production. more Where to Buy Support As cost efficient to run as a mono-only laser printer Paper capacity of 700 sheets from two media sources Easy to operate with advanced printer driver Memory expandable from 32Mb to 1024Mb Pre-configured models available with Wireless 802.11b, Adobe ® PostScript ® Level 3™ and two-sided printing The AcuLaser C1900 is available in 5 configurations: - AcuLaser C1900S: with 32MB, 200 Sheet MP Tray, 10/100BaseTX Networking - AcuLaser C1900: with 32MB, 200 Sheet MP Tray, 500 Sheet Cassette, 10/100BaseTX Networking Support Epson AcuLaser C4100 High performance colour lasers for all your business printing needs The Epson AcuLaser C4100 provides businesses with a high performance colour and monochrome printing solution. It adds crucial colour to your business, while producing high quality monochrome output at lower costs than many monochrome-only printers, and it's just as easy to operate. So now there's no reason to buy two printers, because perfect monochrome and colour solutions are available in one. more Where to Buy Support Epson AcuLaser C9600 Professional high performance A3W colour laser printer Epson AcuLaser C9600 is the perfect professional printing solution for users who require exceptional quality colour and mono output on a range of media formats from C5 up to A3W in size. The Epson AcuLaser C9600 is able to achieve superb print quality by utilising a combination of Epson's exclusive AcuLaser Color Laser Technologies. more Where to Buy Support - AcuLaser C1900PS: with Adobe® PostScript® 3™, 96MB, 200 Sheet MP Tray, 500 Sheet Cassette, 10/100BaseTX Networking - AcuLaser C1900D: with Duplex unit (two sided printing) 96MB, 200 Sheet MP Tray, 500 Sheet Cassette, 10/100BaseTX Networking - AcuLaser C1900 WiFi: with 32MB, 200 Sheet MP Tray, 500 Sheet Cassette, Wireless Networking facility Add colour to your business with the Epson AcuLaser C900 from Epson. Its perfect for the smaller workgroup, being a compact and cost effective laser printing workhorse that offers amazing colour output as well as high Support Epson AcuLaser C4000 High performance colour laser The Epson AcuLaser C4000 provides businesses with high performance colour and monochrome printing solutions more Where to Buy Epson AcuLaser C9100 High speed A3 colour laser printer Why have separate black and white and colour printers when you can have the Epson AcuLaser C9100? Epson has taken the lead in laser technology to deliver a complete high-performance solution for all your colour and mono printing needs. Support EPL-6200L High performance A4 mono laser professional printers The Epson EPL-6200 and EPL-6200L are the ideal printing solutions for small to medium workgroups and personal users. They deliver professional performance quickly, easily, reliably and cost-effectively, and are perfect for users who need high levels of laser quality and productivity at a low investment. more Where to Buy Support EPL-6200 High performance A4 mono laser professional printers The Epson EPL-6200 and EPL-6200L are the ideal printing solutions for small to medium workgroups and personal users. They deliver professional performance quickly, easily, reliably and cost-effectively, and are perfect for users who need high levels of laser quality and productivity at a low investment. more performance black and white production. For the first time, you can now bring the power of high quality colour to your documents without suffering the high costs or low speeds traditionally associated with colour

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.

at first:

```
tyntd-iafhatawiaoihrdemot  lytdws  e ,tfti, astai f ogoh eoase rrranbyne 'nhthnee e
plia tklrgd t o idoe ns,smtt   h ne etie h,hregtrs nigtike,aoaenns lng
```

train more

```
"Tmont thithey" fomesscerliund
Keushey. Thom here
sheulke, anmerenith ol sivh I lalterthend Bleipile shuwy fil on aseterlome
coaniogennc Phe lism thond hon at. MeiDimorotion in ther thize."
```

train more

```
Aftair fall unsuch that the hall for Prince Velzonski's that me of
her hearly, and behs to so arwage fiving were to it beloge, pavu say falling misfort
how, and Gogition is so overelical and ofter.
```

train more

```
"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.
```

PANDARUS:
Alas, I think he shall be come approached and the day
When little srain 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:
I'll drink it.

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.

open source textbook on algebraic geometry

The Stacks Project

home about tags explained tag lookup browse search bibliography recent comments blog add slogans

Browse chapters

Part	Chapter	online	TeX source	view pdf
Preliminaries				
	1. Introduction	online	tex	pdf
	2. Conventions	online	tex	pdf
	3. Set Theory	online	tex	pdf
	4. Categories	online	tex	pdf
	5. Topology	online	tex	pdf
	6. Sheaves on Spaces	online	tex	pdf
	7. Sites and Sheaves	online	tex	pdf
	8. Stacks	online	tex	pdf
	9. Fields	online	tex	pdf
	10. Commutative Algebra	online	tex	pdf

Parts

1. Preliminaries
2. Schemes
3. Topics in Scheme Theory
4. Algebraic Spaces
5. Topics in Geometry
6. Deformation Theory
7. Algebraic Stacks
8. Miscellany

Statistics

The Stacks project now consists of

- 455910 lines of code
- 14221 tags (56 inactive tags)
- 2366 sections

Latex source

For $\bigoplus_{n=1,\ldots,m}$ where $\mathcal{L}_{m_\bullet} = 0$, hence we can find a closed subset $\mathcal{H}$ in $\mathcal{H}$ and any sets $\mathcal{F}$ on X, U is a closed immersion of S, then $U \to T$ is a separated algebraic space.

Proof. Proof of (1). It also start we get

$$S = \mathrm{Spec}(R) = U \times_X U \times_X U$$

and the comparicoly in the fibre product covering we have to prove the lemma generated by $\coprod Z \times_U U \to V$. Consider the maps M along the set of points Sch_{fppf} and $U \to U$ is the fibre category of S in U in Section, ?? and the fact that any U affine, see Morphisms, Lemma ??. Hence we obtain a scheme S and any open subset $W \subset U$ in $Sh(G)$ such that $\mathrm{Spec}(R') \to S$ is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that f_i is of finite presentation over S. We claim that $\mathcal{O}_{X,x}$ is a scheme where $x, x', s'' \in S'$ such that $\mathcal{O}_{X,x'} \to \mathcal{O}'_{X',x'}$ is separated. By Algebra, Lemma ?? we can define a map of complexes $\mathrm{GL}_{S'}(x'/S'')$ and we win. $\square$

To prove study we see that $\mathcal{F}|_U$ is a covering of $\mathcal{X}'$, and $\mathcal{T}_i$ is an object of $\mathcal{F}_{X/S}$ for $i > 0$ and $\mathcal{F}_p$ exists and let $\mathcal{F}_i$ be a presheaf of $\mathcal{O}_X$-modules on $\mathcal{C}$ as a $\mathcal{F}$-module. In particular $\mathcal{F} = U/\mathcal{F}$ we have to show that

$$\widetilde{M}^\bullet = \mathcal{I}^\bullet \otimes_{\mathrm{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1}\mathcal{F})$$

is a unique morphism of algebraic stacks. Note that

$$\text{Arrows} = (Sch/S)^{opp}_{fppf}, (Sch/S)_{fppf}$$

and

$$V = \Gamma(S, \mathcal{O}) \longmapsto (U, \mathrm{Spec}(A))$$

is an open subset of X. Thus U is affine. This is a continuous map of X is the inverse, the groupoid scheme S.

Proof. See discussion of sheaves of sets. $\square$

The result for prove any open covering follows from the less of Example ??. It may replace S by $X_{spaces,étale}$ which gives an open subspace of X and T equal to S_{Zar}, see Descent, Lemma ??. Namely, by Lemma ?? we see that R is geometrically regular over S.

Lemma 0.1. *Assume (3) and (3) by the construction in the description.*

Suppose $X = \lim |X|$ (by the formal open covering X and a single map $\underline{Proj}_X(\mathcal{A}) = \mathrm{Spec}(B)$ over U compatible with the complex

$$Set(\mathcal{A}) = \Gamma(X, \mathcal{O}_{X,\mathcal{O}_X}).$$

When in this case of to show that $\mathcal{Q} \to \mathcal{C}_{Z/X}$ 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 the fact that the following theorem

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

Proof. This is form all sheaves of sheaves on X. But given a scheme U and a surjective étale morphism $U \to X$. Let $U \cap U = \coprod_{i=1,\ldots,n} U_i$ be the scheme X over S at the schemes $X_i \to X$ and $U = \lim_i X_i$. $\square$

The following lemma surjective restrocomposes of this implies that $\mathcal{F}_{x_0} = \mathcal{F}_{x_0} = \mathcal{F}_{\mathcal{X},\ldots,0}$.

Lemma 0.2. *Let X be a locally Noetherian scheme over S, $E = \mathcal{F}_{X/S}$. Set $\mathcal{I} = \mathcal{J}_1 \subset \mathcal{I}'_n$. Since $\mathcal{I}^n \subset \mathcal{I}^n$ are nonzero over $i_0 \leq \mathfrak{p}$ is a subset of $\mathcal{J}_{n,0} \circ \overline{A}_2$ works.*

Lemma 0.3. *In Situation ??. Hence we may assume $\mathfrak{q}' = 0$.*

Proof. We will use the property we see that $\mathfrak{p}$ is the mext 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 δ_{n+1} is a scheme over S. $\square$

Proof. Omitted. □

Lemma 0.1. *Let $\mathcal{C}$ be a set of the construction.*

Let $\mathcal{C}$ 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 $X_{\acute{e}tale}$ we have

$$\mathcal{O}_X(\mathcal{F}) = \{morph_1 \times_{\mathcal{O}_X} (\mathcal{G}, \mathcal{F})\}$$

where $\mathcal{G}$ 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 $\mathcal{U} \subset \mathcal{X}$ 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

$$b : X \to Y' \to Y \to Y \to Y' \times_X Y \to X.$$

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 $x \in \mathcal{G}$ the diagram

$$\begin{tikzcd}
S \arrow[r] \arrow[d] & & & \\
\xi \arrow[r] & \mathcal{O}_{X'} \arrow[dr] & & \\
\mathrm{gor}_s \arrow[u] & & & \\
 & = \alpha' \arrow[r] \arrow[d, leftrightarrow] & & \\
 & = \alpha' \arrow[r] & \alpha & X \arrow[d] \\
 & \mathrm{Spec}(K_\psi) & & \mathrm{Mor}_{Sets} \quad \mathrm{d}(\mathcal{O}_{X_{X/k}}, \mathcal{G})
\end{tikzcd}$$

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

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

□

Proof. We have see that $X = \mathrm{Spec}(R)$ 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 $\mathcal{G}$ is a finite presentation, see Lemmas ??.
A *reduced above* we conclude that U is an open covering of $\mathcal{C}$. The functor $\mathcal{F}$ is a "field

$$\mathcal{O}_{X,x} \longrightarrow \mathcal{F}_{\overline{x}} \quad \text{-1}(\mathcal{O}_{X_{\acute{e}tale}}) \longrightarrow \mathcal{O}_{X_i}^{-1}\mathcal{O}_{X_\lambda}(\mathcal{O}_{X_\eta}^{\overline{\eta}})$$

is an isomorphism of covering of $\mathcal{O}_{X_i}$. 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 opens of finite type over S.
If $\mathcal{F}$ is a scheme theoretic image points. □

If $\mathcal{F}$ is a finite direct sum $\mathcal{O}_{X_\lambda}$ is a closed immersion, see Lemma ??. This is a sequence of $\mathcal{F}$ is a similar morphism.

This repository Search

Explore Gist Blog Help

karpathy

torvalds / linux

Watch 3,711 Star 23,054 Fork 9,141

Linux kernel source tree

520,037 commits 1 branch 420 releases 5,039 contributors

branch: master linux / +

Merge branch 'drm-fixes' of git://people.freedesktop.org/~airlied/linux

torvalds authored 9 hours ago latest commit 4b1706927d

Documentation	Merge git://git.kernel.org/pub/scm/linux/kernel/git/nab/target-pending	6 days ago
arch	Merge branch 'x86-urgent-for-linus' of git://git.kernel.org/pub/scm/l...	a day ago
block	block: discard bdi_unregister() in favour of bdi_destroy()	9 days ago
crypto	Merge git://git.kernel.org/pub/scm/linux/kernel/git/herbert/crypto-2.6	10 days ago
drivers	Merge branch 'drm-fixes' of git://people.freedesktop.org/~airlied/linux	9 hours ago
firmware	firmware/ihex2fw.c: restore missing default in switch statement	2 months ago
fs	vfs: read file_handle only once in handle_to_path	4 days ago
include	Merge branch 'perf-urgent-for-linus' of git://git.kernel.org/pub/scm/...	a day ago
init	init: fix regression by supporting devices with major:minor:offset fo...	a month ago

Code

Pull requests 74

Pulse

Graphs

HTTPS clone URL

https://github.c

You can clone with HTTPS, SSH, or Subversion.

Clone in Desktop

Download ZIP

```c
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 << 1))
      pipe = (in_use & UMXTHREAD_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 ");
}
```

Generated C code

```
/*
 *  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/setew.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));   \
  if (__type & DO_READ)

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;
}
```

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Recurrent Neural Network

Recurrent Neural Network

usually want to predict
a vector at some time steps

Recurrent Neural Network

We can process a sequence of vectors **x** by applying a recurrence formula at every time step:

$$\boxed{h_t} = \boxed{f_W}(\boxed{h_{t-1}}, \boxed{x_t})$$

new state

old state

input vector at some time step

some function with parameters W

Recurrent Neural Network

We can process a sequence of vectors **x** by applying a recurrence formula at every time step:

$$h_t = f_W(h_{t-1}, x_t)$$

Notice: the same function and the same set of parameters are used at every time step.

(Vanilla) Recurrent Neural Network

The state consists of a single *"hidden"* vector **h**:

$$h_t = f_W(h_{t-1}, x_t)$$

$$h_t = \tanh(W_{hh} h_{t-1} + W_{xh} x_t)$$

$$y_t = W_{hy} h_t$$

Backpropagation through time

Forward through entire sequence to compute loss, then backward through entire sequence to compute gradient

Long Short Term Memory (LSTM)

Vanilla RNN

$$h_t = \tanh\left(W\begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}\right)$$

LSTM

$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} W \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}$$

$$c_t = f \odot c_{t-1} + i \odot g$$

$$h_t = o \odot \tanh(c_t)$$

Hochreiter and Schmidhuber, “Long Short Term Memory”, Neural Computation 1997

Understanding LSTMs: colah.github.io/posts/2015-08-Understanding-LSTMs

GRU [*Learning phrase representations using rnn encoder-decoder for statistical machine translation*, Cho et al. 2014]

$$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r)$$
$$z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z)$$
$$\tilde{h}_t = \tanh(W_{xh}x_t + W_{hh}(r_t \odot h_{t-1}) + b_h)$$
$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t$$

[*LSTM: A Search Space Odyssey*, Greff et al., 2015]

[*An Empirical Exploration of Recurrent Network Architectures,* Jozefowicz et al., 2015]

MUT1:

$$\begin{aligned} z &= \mathrm{sigm}(W_{\mathrm{xz}}x_t + b_{\mathrm{z}}) \\ r &= \mathrm{sigm}(W_{\mathrm{xr}}x_t + W_{\mathrm{hr}}h_t + b_{\mathrm{r}}) \\ h_{t+1} &= \tanh(W_{\mathrm{hh}}(r \odot h_t) + \tanh(x_t) + b_{\mathrm{h}}) \odot z \\ &+ h_t \odot (1 - z) \end{aligned}$$

MUT2:

$$\begin{aligned} z &= \mathrm{sigm}(W_{\mathrm{xz}}x_t + W_{\mathrm{hz}}h_t + b_{\mathrm{z}}) \\ r &= \mathrm{sigm}(x_t + W_{\mathrm{hr}}h_t + b_{\mathrm{r}}) \\ h_{t+1} &= \tanh(W_{\mathrm{hh}}(r \odot h_t) + W_{xh}x_t + b_{\mathrm{h}}) \odot z \\ &+ h_t \odot (1 - z) \end{aligned}$$

MUT3:

$$\begin{aligned} z &= \mathrm{sigm}(W_{\mathrm{xz}}x_t + W_{\mathrm{hz}}\tanh(h_t) + b_{\mathrm{z}}) \\ r &= \mathrm{sigm}(W_{\mathrm{xr}}x_t + W_{\mathrm{hr}}h_t + b_{\mathrm{r}}) \\ h_{t+1} &= \tanh(W_{\mathrm{hh}}(r \odot h_t) + W_{xh}x_t + b_{\mathrm{h}}) \odot z \\ &+ h_t \odot (1 - z) \end{aligned}$$

VIDEO PROCESSING

Semi-automatic Data Collection for Understanding Video with Fine-grained Motion (Baseball Video Dataset)

Yonsei University

Seon Joo Kim

Sports Highlights

Video Summarization / Highlight Detection

[1]

[2]

[1] W. Chu et. al., “Video Co-summarization: Video Summarization by Visual Co-occurrence”, CVPR 2015
[2] T. Yao et. al., “Highlight Detection with Pairwise Deep Ranking for First-Person Video Summarization”, CVPR 2016

Query-based Highlights for Baseball Games

Input Video (sports broadcast)

Question

① Show me the scoring footages of team A from the game last night

② Show me all the homeruns from all pro baseball games from yesterday

user

Generate Question-based Highlights

① Scoring highlights of team A

② Homerun collection highlights

Video Recognition, Video Highlight/Summarization, VQA, NLP

Query-based Highlights for Baseball Games

Understanding Baseball Games: Automatic Play-by-Play Generation

All Plays | Scoring Plays

Dodgers - Top 1st

SIMS PITCHING FOR ATL

Taylor flied out to center.

PITCH		TYPE	MPH
1	Strike Looking	Fastball	91
2	Strike Swinging	Fastball	92
3	Ball	Curve	73
4	Foul Ball	Slider	86
5	Ball	Fastball	92
6	Fly Out	Slider	88

Seager popped out to shortstop.

Turner singled to right.

Bellinger lined out to center.

Dataset???

Baseball Video Database (BBDB)

1,172 games

4,254 hours of baseball

500 million frames

30 activity labels

Per game, we also have:

- Play-by-play text
- Highlight

Baseball Video Database (BBDB)

But how do label all those videos???

Semi-Automatic Labelling

Data Collection

vs. Other Video Dataset

Dataset	#instances†/#videos	Avg. Duration	Untrimmed	Detection	Sequential‡
UCF101 [32]	/13k	7s	-	-	-
HMDB51 [21]	/7k	3s	-	-	-
Kinetics [4]	/306k	10s	-	-	-
Sports1M [17]	/1.1m	300s	✓	-	-
Youtube8M [1]	/8.3m	230s	✓	-	-
THUMOS15 [13]	23.1k/21k	4s	✓	✓	△
MultiTHUMOS [38]	39k/400	270s	✓	✓	△
ActivityNet [3]	28k/20k	180s	✓	✓	△
Hollywood2 [23]	-/3.7k	20s	✓	✓	△
Cooking [27]	5.6k/44	600s	✓	✓	✓
Breakfast [20]	11k/2k	140s	✓	✓	✓
BBDB (ours)	**405k**/1k	**13,000s**	✓	✓	✓

IoU threshold	0.3	0.4	0.5	0.6	0.7
BBDB	98.2	95.5	89.0	68.7	49.5

Table 1. Precision of semi-automatic labeling against human labeling on the BBDB. IoU threshold ranges from 0.3 to 0.7.

Class	count	mean	stdev
Foul fly out	8	9.16	0.96
Sacrifice bunt out	2	8.86	1.00
Double play	13	8.36	1.34
Deadball	23	6.61	1.36
Strike	333	5.45	1.44
Swing and a miss	141	5.45	1.44
⋮			
Wild pitch	6	7.50	4.12
Error	11	12.90	4.87
Home in	76	14.93	6.18
Homerun	16	17.38	8.97

Table 2. Clip length statistics on 6 manually annotated games. Count is the number of clips, and the unit of mean is second.

Algorithm Evaluation

Method	mAP	Cost
Action Recognition		
IDT + FV [36]	23.6	12 days
Single frame [31]	35.0	2 days
Optical flow stacking [30]	34.1	1 day
Two-stream (Avg) [30]	36.9	3 days
Two-stream (SVM) [30]	41.7	3 days
C3D [35]	40.2	18 hrs
CNN+GRU [8] (Oversampling)	**62.8**	2.5 days

Ball (True Positive)

Ground Truth : Ball
Predicted action (probability) : Ball (0.998)

Ground Truth : Ball
Predicted action (probability) : Ball (0.999)

Ball (False Positive)

Ground Truth : Strike
Predicted action (probability) : Ball (0.999)

Ground Truth : Strike
Predicted action (probability) : Ball (0.696)

Hit by Pitch (True Positive)

Ground Truth : Hit by Pitch
Predicted action (probability) : Hit by Pitch (1.0)

Ground Truth : Hit by Pitch
Predicted action (probability) : Hit by Pitch (1.0)

Hit by Pitch (False Positive)

Ground Truth : Foul
Predicted action (probability) : Hit by Pitch (0.999)

Ground Truth : Foul
Predicted action (probability) : Hit by Pitch (0.885)

What is DNN looking at?

K. Simonyan et al., “Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps”, ECCV 2014

What is DNN looking at?

What is DNN looking at?

What is DNN looking at?

Summary

- Large database labelled semi-automatically
- Good for fine-grained motion recognition
- Can work on many tasks:

  - video recognition

  - video temporal localization

  - highlight generation

  - data imbalance problem

  - may be even more…

Deep Learning for Videos

ConvNets for images

Feature-based approaches to Activity Recognition

Dense trajectories and motion boundary descriptors for action recognition
Wang et al., 2013

Action Recognition with Improved Trajectories
Wang and Schmid, 2013

(code available!)

Dense trajectories

Dense trajectories and motion boundary descriptors for action recognition

Wang et al., 2013

detect feature points

track features with optical flow

extract HOG/HOF/MBH features in the (stabilized) coordinate system of each tracklet

Dense trajectories and motion boundary descriptors for action recognition

Wang et al., 2013

detected feature points

[J. Shi and C. Tomasi, “Good features to track,” CVPR 1994]
[Ivan Laptev 2005]

Dense trajectories and motion boundary descriptors for action recognition

Wang et al., 2013

track each keypoint using **optical flow.**

[G. Farnebäck, “Two-frame motion estimation based on polynomial expansion,” 2003]
[T. Brox and J. Malik, “Large displacement optical flow: Descriptor matching in variational motion estimation,” 2011]

Dense trajectories and motion boundary descriptors for action recognition
Wang et al., 2013

Extract features in the local coordinate system of each tracklet.

Accumulate into histograms, separately acco rding to multiple spatio-temporal layouts.

Case Study: AlexNet

[Krizhevsky et al. 2012]

Input: 227x227x3 images

First layer (CONV1): 96 11x11 filters applied at stride 4
=>
Output volume **[55x55x96]**
Q: What if the input is now a small chunk of video? E.g. [227x227x3x15] ?

Case Study: AlexNet

[Krizhevsky et al. 2012]

Input: 227x227x3 images

First layer (CONV1): 96 11x11 filters applied at stride 4
=>
Output volume **[55x55x96]**
Q: What if the input is now a small chunk of video? E.g. [227x227x3x15] ?
A: Extend the convolutional filters in time, perform spatio-temporal convolutions!
E.g. can have 11x11xT filters, where T = 2..15.

Spatio-Temporal ConvNets

Figure 3. A 3D CNN architecture for human action recognition. This architecture consists of 1 hardwired layer, 3 convolution layers, 2 subsampling layers, and 1 full connection layer. Detailed descriptions are given in the text.

[3D Convolutional Neural Networks for Human Action Recognition, Ji et al., 2010]

Spatio-Temporal ConvNets

Sequential Deep Learning for Human Action Recognition, Baccouche et al., 2011

Spatio-Temporal ConvNets

spatio-temporal convolutions; worked best.

[Large-scale Video Classification with Convolutional Neural Networks, Karpathy et al., 2014]

Spatio-Temporal ConvNets

Learned filters on the first layer

[Large-scale Video Classification with Convolutional Neural Networks, Karpathy et al., 2014]

Spatio-Temporal ConvNets

1 million videos
487 sports classes

[Large-scale Video Classification with Convolutional Neural Networks, Karpathy et al., 2014]

Spatio-Temporal ConvNets

Model	Clip Hit@1	Video Hit@1	Video Hit@5
Feature Histograms + Neural Net	-	55.3	-
Single-Frame	41.1	59.3	77.7
Single-Frame + Multires	**42.4**	**60.0**	**78.5**
Single-Frame Fovea Only	30.0	49.9	72.8
Single-Frame Context Only	38.1	56.0	77.2
Early Fusion	38.9	57.7	76.8
Late Fusion	40.7	59.3	78.7
Slow Fusion	**41.9**	60.9	**80.2**
CNN Average (Single+Early+Late+Slow)	41.4	63.9	82.4

The motion information didn't add all that much...

[Large-scale Video Classification with Convolutional Neural Networks, Karpathy et al., 2014]

Spatio-Temporal ConvNets

Spatio-Temporal ConvNets

Figure 3. **C3D architecture**. C3D net has 8 convolution, 5 max-pooling, and 2 fully connected layers, followed by a softmax output layer. All 3D convolution kernels are $3 \times 3 \times 3$ with stride 1 in both spatial and temporal dimensions. Number of filters are denoted in each box. The 3D pooling layers are denoted from `pool1` to `pool5`. All pooling kernels are $2 \times 2 \times 2$, except for `pool1` is $1 \times 2 \times 2$. Each fully connected layer has 4096 output units.

3D VGGNet, basically.

[Learning Spatiotemporal Features with 3D Convolutional Networks, Tran et al. 2015]

Spatio-Temporal ConvNets

(of VGGNet fame)

[Two-Stream Convolutional Networks for Action Recognition in Videos, **Simonyan** and Zisserman 2014]

[T. Brox and J. Malik, “Large displacement optical flow: Descriptor matching in variational motion estimation,” 2011]

Spatio-Temporal ConvNets

Spatial stream ConvNet	73.0%	40.5%
Temporal stream ConvNet	83.7%	54.6%
Two-stream model (fusion by averaging)	86.9%	58.0%
Two-stream model (fusion by SVM)	**88.0%**	**59.4%**

Two-stream version works much better than either alone.

[Two-Stream Convolutional Networks for Action Recognition in Videos, **Simonyan** and Zisserman 2014]

[T. Brox and J. Malik, “Large displacement optical flow: Descriptor matching in variational motion estimation,” 2011]

Long-time Spatio-Temporal ConvNets

All 3D ConvNets so far used local motion cues to get extra accuracy (e.g. half a second or so)

Q: what if the temporal dependencies of interest are much much longer? E.g. several seconds?

Long-time Spatio-Temporal ConvNets

(This paper was way ahead of its time. Cited 65 times.)

Sequential Deep Learning for Human Action Recognition, Baccouche et al., **2011**

Long-time Spatio-Temporal ConvNets

LSTM way before it was cool

(This paper was way ahead of its time. Cited 65 times.)

Sequential Deep Learning for Human Action Recognition, Baccouche et al., **2011**

Long-time Spatio-Temporal ConvNets

[Long-term Recurrent Convolutional Networks for Visual Recognition and Description, Donahue et al., 2015]

Long-time Spatio-Temporal ConvNets

[Beyond Short Snippets: Deep Networks for Video Classification, Ng et al., 2015]

Summary so far

We looked at two types of architectural patterns:

1. Model temporal motion locally (3D CONV)

2. Model temporal motion globally (LSTM / RNN)

\+ Fusions of both approaches at the same time.

Summary so far

We looked at two types of architectural patterns:

1. Model temporal motion locally (3D CONV)

2. Model temporal motion globally (LSTM / RNN)

\+ Fusions of both approaches at the same time.

There is another (cleaner) way!

RNN
Infinite (in theory) temporal extent
(neurons that are function of all video frames in the past)
3D C ONVNET
Finite temporal extent
(neurons that are only a function of finitely many video frames in the past)
video

Long-time Spatio-Temporal ConvNets

Beautiful:
All neurons in the ConvNet are recurrent.

$$
\begin{aligned}
\mathbf{z}_t^l &= \sigma(\mathbf{W}_z^l * \mathbf{x}_t^l + \mathbf{U}_z^l * \mathbf{h}_{t-1}^l), \\
\mathbf{r}_t^l &= \sigma(\mathbf{W}_r^l * \mathbf{x}_t^l + \mathbf{U}_r^l * \mathbf{h}_{t-1}^l), \\
\tilde{\mathbf{h}}_t^l &= \tanh(\mathbf{W}^l * \mathbf{x}_t^l + \mathbf{U} * (\mathbf{r}_t^l \odot \mathbf{h}_{t-1}^l), \\
\mathbf{h}_t^l &= (1 - \mathbf{z}_t^l)\mathbf{h}_{t-1}^l + \mathbf{z}_t^l \tilde{\mathbf{h}}_t^l,
\end{aligned}
$$

Only requires (existing) 2D CONV routines. No need for 3D spatio-temporal CONV.

[Delving Deeper into Convolutional Networks for Learning Video Representations, Ballas et al., 2016]

Long-time Spatio-Temporal ConvNets

Normal ConvNet:

[Delving Deeper into Convolutional Networks for Learning Video Representations, Ballas et al., 2016]

Long-time Spatio-Temporal ConvNets

[Delving Deeper into Convolutional Networks for Learning Video Representations, Ballas et al., 2016]

Long-time Spatio-Temporal ConvNets

Recall: RNNs

$$h_t = f_W(h_{t-1}, x_t)$$

Vanilla RNN

$$h_t = \tanh(W_{hh} h_{t-1} + W_{xh} x_t)$$

GRU

$$\begin{aligned}
\mathbf{z}_t &= \sigma(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1}), \\
\mathbf{r}_t &= \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{U}_r \mathbf{h}_{t-1}), \\
\tilde{\mathbf{h}}_t &= \tanh(\mathbf{W}\mathbf{x}_t + \mathbf{U}(\mathbf{r}_t \odot \mathbf{h}_{t-1}) \\
\mathbf{h}_t &= (1 - \mathbf{z}_t)\mathbf{h}_{t-1} + \mathbf{z}_t \tilde{\mathbf{h}}_t,
\end{aligned}$$

LSTM

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

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

$$h_t^l = o \odot \tanh(c_t^l)$$

[Delving Deeper into Convolutional Networks for Learning Video Representations, Ballas et al., 2016]

Long-time Spatio-Temporal ConvNets

Recall: RNNs

$$h_t = f_W(h_{t-1}, x_t)$$

Matrix multiply
=> C
ONV

GRU

$$\begin{aligned}
\mathbf{z}_t &= \sigma(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1}), \\
\mathbf{r}_t &= \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{U}_r \mathbf{h}_{t-1}), \\
\tilde{\mathbf{h}}_t &= \tanh(\mathbf{W}\mathbf{x}_t + \mathbf{U}(\mathbf{r}_t \odot \mathbf{h}_{t-1}) \\
\mathbf{h}_t &= (1 - \mathbf{z}_t)\mathbf{h}_{t-1} + \mathbf{z}_t \tilde{\mathbf{h}}_t,
\end{aligned}$$

$$\begin{aligned}
\mathbf{z}_t^l &= \sigma(\mathbf{W}_z^l * \mathbf{x}_t^l + \mathbf{U}_z^l * \mathbf{h}_{t-1}^l), \\
\mathbf{r}_t^l &= \sigma(\mathbf{W}_r^l * \mathbf{x}_t^l + \mathbf{U}_r^l * \mathbf{h}_{t-1}^l), \\
\tilde{\mathbf{h}}_t^l &= \tanh(\mathbf{W}^l * \mathbf{x}_t^l + \mathbf{U} * (\mathbf{r}_t^l \odot \mathbf{h}_{t-1}^l), \\
\mathbf{h}_t^l &= (1 - \mathbf{z}_t^l)\mathbf{h}_{t-1}^l + \mathbf{z}_t^l \tilde{\mathbf{h}}_t^l,
\end{aligned}$$

[Delving Deeper into Convolutional Networks for Learning Video Representations, Ballas et al., 2016]

RNN
Infinite (in theory) temporal extent
(neurons that are function of all video frames in the past)
3D C ONVNET
Finite temporal extent
(neurons that are only a function of finitely many video frames in the past)
video

i.e. we obtain:

Infinite (in theory)
temporal extent

(neurons that are function
of all video frames in the past)

Summary

- You think you need a Spatio-Temporal Fancy Video ConvNet
- STOP. Do you really?
- Okay fine: do you want to model:
  - local motion? (use 3D CONV), or
  - global motion? (use LSTM).
- Try out using Optical Flow in a second stream (can work better sometimes)
- Try out GRU-RCN! (imo best model)

Deep Video Super-Resolution Network Using Dynamic Upsampling Filters Without Explicit Motion Compensation

Younghyun Jo Seoung Wug Oh Jaeyeon Kang Seon Joo Kim

Yonsei University

CVPR18

Super-Resolution (SR)

Medical imaging

Satellite imaging

Surveillance camera

8K UHD TV

[CES 2018] 삼성전자, 85인치 QLED TV 공개…"AI 기술로 저화질→고화질"

2018. 01. 07.

Video SR

- Conventional video SR algorithm consists of two steps.
  - Motion compensation (MC) followed by restoration process.

Set of low-res (LR) frames Motion compensated Reconstructed high-res (HR) frame

- But it has **a problem**.
  - <u>SR results rely heavily on the accurate motion estimation.</u>

Input LR frames The result of inaccurate(d) and accurate MC(e)

[Ma et al. CVPR 2015]

✓ We make our SR network implicitly utilized the motion information.

To Avoid Explicit MC

- ✓ We propose dynamic upsampling filters.
  - Upsampling filters for each pixel are generated depending on spatiotemporal neighborhood.

Network Architecture

- The network implicitly learns the motion information and utilizes it to generate the filters.

Dynamic Upsampling Filters

An example of upscaling a pixel at location (3,3) in the center input frame X_t by the upscaling factor $r = 4$. 16 generated filters from $F_t^{3,3,0,0}$ to $F_t^{3,3,3,3}$ are used to create 16 pixels in the region (12,12) to (15,15) of the filtered HR frame $\tilde{Y}_t$.

Datasets

- Training dataset.
  - 351 videos including wildlife, activity, and landscape, which contain various real-world motions and textures.
  - Total 160,000 training data with sufficient amount of motion is extracted from them.
- Data augmentation to the temporal axis.
  - Reverse order, sampling interval.

- Test dat

Synthetic Motion Test

Synthetic Motion Test

Comparisons with Other Methods

Upscale	Metric	Bicubic	VSRnet	VESPCN	Tao *et al.*	Liu *et al.*	Ours-10L	Ours-16L	Ours-28L
×2	PSNR	28.43	31.30	-	-	-	**33.73**	-	-
	SSIM	0.8676	0.9278	-	-	-	**0.9554**	-	-
×3	PSNR	25.28	26.79	27.25	-	-	**28.90**	-	-
	SSIM	0.7329	0.8098	0.8447	-	-	**0.8898**	-	-
×4	PSNR	23.79	24.84	25.35	26.01	25.24	**26.81**	**26.99**	**27.34**
	SSIM	0.6332	0.7049	0.7557	0.7744	-	**0.8145**	**0.8215**	**0.8327**

Result Video

Qualitative Comparisons

Summary

- **No explicit motion compensation.**
  - Fundamentally different framework for video SR.
  - The motion information is implicitly utilized to generate dynamic upsampling filters instead.
- **Dynamic upsampling filters.**
  - They are generated locally and dynamically depending on spatiotemporal neighborhood of each pixel.
  - HR frame is directly reconstructed by local filtering to the input center frame.
- **Residual learning.**
  - The filtering alone lacks sharpness as it is still a linear operation.
  - To enhance high frequency details.
- **New data augmentation.**
  - To the temporal axis as well as to the spatial axis.

Fast Video Object Segmentation by Reference Guided Mask Propagation

Seoung Wug Oh, Joon-Young Lee, Kalyan Sunkavalli, Seon Joo Kim

CVPR 2018

© 2017 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

Semi-supervised Video Object Segmentation

Only the first frame annotation is given

© 2017 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

Related Works: deep learning based video segmentation

Detection-based methods

- **Strong:** drift and occlusion
- **Weak:** appearance changes

OSVOS
[Caelles et al, CVPR'17]

Propagation-based methods

- **Strong:** appearance changes
- **Weak:** drift and occlusions

MaskTrack
[Perazzi and Khoreva et al, CVPR'17]

Online learning

- Fine-tune network on the target object that appears on the first frame
- **Slow**

© 2017 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

Our Approach

Using a **single model,** combine strengths of both lines of approaches:

- Appearance of the object (Detection-based)
- Temporal consistency (Propagation-based)

Avoid Online Learning

- Perform object matching **at running time**

Siamese Encoder-Decoder network

- Previous frame's mask (to propagate)
- Reference frame (to detect)

© 2017 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

Siamese Encoder-Decoder network

Inputs

- **(Target stream)**
  - Target frame image
  - previous frame's mask
- **(Reference stream)**
  - Reference frame image
  - it's ground truth mask

Output

- Target frame mask

Target frame image + previous frame's mask

Our network

Target frame mask

Reference frame image + It's GT mask

Frames and masks are shown overlaid

© 2017 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

Siamese Encoder-Decoder network

ResNet-50 [He et al, CVPR'16]

[Peng et al, arXiv'17]

[Pinheiro et al, ECCV'16]

© 2017 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

Fine-tune on videos: DAVIS-2017

Training with recurrence: mask output is repeatedly used as input for the next estimation.

Back-propagation through time (like RNN)

© 2017 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

Two-stage Training

No video segmentation data that provide enough training data

- Recent DAVIS dataset provide 60 short videos for training [Perazzi and J. Pont-Tuset]

First, pre-train using static images

- By **simulating** our network's inputs

Second, fine-tune on videos

© 2017 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

Pre-training on static images

Datasets: Pascal VOC, ECSSD, MSRA10K

Generation Method 1:

Random transform

- Realistic but not diverse

Target frame

Reference frame

© 2017 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

Pre-training on static images

Datasets: Pascal VOC, ECSSD, MSRA10

Generation Method 1:

Random transform

- Realistic but not diverse

Generation Method 2:

Image composition

- Diverse, but unrealistic

© 2017 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

Results

No Online learning
No Post-processing

Quantitative results on DAVIS

		VPN	MSK	LCT	PLM	OSVOS	OnAVOS	SegFlow	Ours
Single Object	IoU	70.2	79.7	80.5	70.0	79.8	**86.1**	74.8	**81.5**
	F	65.5	75.4	77.6	62.0	80.6	**84.9**	74.5	**82.0**
Multi Object	IoU	-	-	-	-	52.1	**61.0**	-	**64.8**
	F	-	-	-	-	-	**66.1**	-	**68.6**

Single Object: DAVIS-2016 validation

Multiple Object: DAVIS-2017 validation

1st **2nt**

© 2017 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

Results

No Online learning
No Post-processing

Quantitative results on DAVIS

Single Object: DAVIS-2016 validation

© 2017 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

Results: Single Object

© 2017 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

Results: Single Object

© 2017 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

Results: Multiple Object

© 2017 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

GEOMETRY

Image alignment

Two-view geometry

Triangulation: Geometric approach

- Find shortest segment connecting the two viewing rays and let X be the midpoint of that segment

Epipolar geometry

- **Baseline** – line connecting the two camera centers
- **Epipolar Plane** – plane containing baseline
- **Epipoles**

= intersections of baseline with image planes
= projections of the other camera center

Binocular stereo

- Given a calibrated binocular stereo pair, fuse it to produce a depth image

image 1

image 2

Dense depth map

Beyond two-view stereo

The third view can be used for verification

Stereo from community photo collections

M. Goesele, N. Snavely, B. Curless, H. Hoppe, S. Seitz, Multi-View Stereo for Community Photo Collections, ICCV 2007

http://grail.cs.washington.edu/projects/mvscpc/

Stereo from community photo collections

Comparison: 90% of points within 0.128 m of laser scan (building height 51m)

M. Goesele, N. Snavely, B. Curless, H. Hoppe, S. Seitz, Multi-View Stereo for Community Photo Collections, ICCV 2007

Recent Advances in Geometry in Vision

http://www.youtube.com/watch?v=36PFT6SkYMI

EPINET: A Fully-Convolutional Neural Network using Epipolar Geometry for Depth from Light Field Images

CVPR2018

Changha Shin, Hae Gon Jeon, Young Jin Yoon, In-So Kweon, Seon Joo Kim

Yonsei University

Which viewpoints are selected to estimate disparity map?

- Total N x N viewpoints

"The Variational Structure of Disparity and Regularization of 4D Light Fields."
B. Goldluecke, S. Wanner *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

"Depth from combining defocus and correspondence using light-field cameras."
Tao, Michael W., et al**.** *2013 IEEE International Conference on*. IEEE, *Computer Vision*

"Accurate Depth Map Estimation from a Lenslet Light Field Camera."
Hae-Gon Jeon., et al**.** *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

Our Network Architecture

- Design multi-stream networks which can preserve the **Epipolar Property** of light field image.

Table 1. The effect of the number of viewpoints on performance.

	1-stream	2-streams	4-streams
Input Views			
MSE	2.165	1.729	**1.393**
Bad pixel ratio (<0.07px)	7.61	5.94	**3.87**

Data Augmentation Techniques for LF images

- View-shift augmentation

- Scale augmentation

- Rotation augmentation

Augular resolution	3 × 3	5 × 5	7 × 7					9 × 9
Augmentaion type	Full Aug	Full Aug	Color	Color + Viewshift	Color + Rotation	Color + scaling	Full Aug	Full Aug
Mean square error	1.568	1.475	2.799	2.564	1.685	2.33	1.434	1.461
Bad pixel ratio (>0.07px)	8.63	4.96	6.67	6.29	5.54	5.69	3.94	3.91

Results

- 4D Light Field Benchmark
  - 16 synthetic light-field images(9x9 viewpoints)
  - Depth/disparity map for training scenes
  - Site: http://hci-lightfield.iwr.uni-heidelberg.de

Comparison of the accuracy and the runtime of light field depth estimation algorithms.

Bad pixel (Error<0.03)

Algorithm	Meta MEDIAN No preview		Meta AVG No preview	
Epinet-fcn-m	7.731	1	9.537	1
Epinet-fcn	9.501	3	10.745	2
Epinet-fcn9x9	9.058	2	11.212	3
OBER-cross+ANP	11.018	4	13.100	4
SPO-MO	15.243	5	14.258	5
OBER-cross	15.465	6	18.731	6
OFSY_330/DNR	20.373	8	20.225	7
PS_RF	19.719	7	21.630	8
SPO	25.215	15	22.300	9
GLFCV	23.479	11	22.450	10
ZCTV	25.721	16	22.917	11
CAE	23.386	10	22.949	12
RM3DE	23.561	12	23.259	13
OBER	23.713	13	23.565	14
EPN+OS+GC	21.731	9	23.828	15

Mean Square Error (multiplied with 100)

Algorithm	Meta MEDIAN No preview		Meta AVG No preview	
Epinet-fcn-m	1.203	1	2.418	1
Epinet-fcn	1.208	2	2.476	2
Epinet-fcn9x9	1.280	3	2.521	3
OBER-cross+ANP	1.464	5	2.584	4
SPO-MO	1.805	7	3.518	5
CAE	2.667	11	3.730	6
FBS*	1.701	6	3.805	7
RM3DE	1.455	4	3.922	8
SPO	3.309	15	3.968	9
OBER-cross	2.547	10	4.010	10
OBER	2.381	9	4.616	11
PS_RF	2.169	8	4.617	12
RPRF	3.760	17	5.683	13
RPRF-5view	3.295	14	5.948	14
*EPI1	3.932	18	5.975	15

Runtime (seconds)

Algorithm	Meta MEDIAN No preview		Meta AVG No preview		Algorithm	Meta MEDIAN No preview		Meta AVG No preview	
MVCMv0	0.110	1	0.111	1	*EPI1	85.045	15	88.194	15
Epinet-fcn	1.972	2	1.976	2	OBER-cross	96.200	16	96.417	16
Epinet-fcn9x9	2.032	3	2.041	3	OBER-cross+ANP	174.595	17	182.994	17
GLFCV	4.049	4	4.050	4	OFSY_330/DNR	198.299	18	200.282	18
BSL_I	4.640	5	5.456	5	EPN+OS+GC	260.602	19	274.710	19
BSL	5.149	6	5.962	6	LF_OCC	484.883	20	519.868	20
*EPI2	8.789	7	8.406	7	FBS*	680.850	21	655.905	21
Epinet-fcn-m	10.647	8	10.659	8	CAE	822.272	22	832.081	22
RPRF-5view	12.966	9	12.498	9	*LF	994.311	23	1009.756	23
RPRF	35.456	10	34.529	10	OMG_occ	1049.401	24	1163.938	24
ZCTV	37.400	11	37.450	11	PS_RF	1141.815	25	1412.623	25
OBER	40.420	12	40.141	12	SC_GC	2052.190	26	2056.344	26
*MV	45.638	14	40.879	13	SPO	2111.500	27	2115.417	27
RM3DE	45.149	13	47.434	14	SPO-MO	4297.000	28	4304.333	28

Synthetic Results

Real-world Results

B: Globally consistent depth labeling of 4D lightfields. S. Wanner and B. Goldluecke
C: Accurate depth map estimation from a lenslet light field camera.H.-G. Jeon,
D: Robust light field depth estimation for noisy scene with occlusion. W. Williem
E: Occlusionaware depth estimation using light-field cameras. T.-C. Wang,.
F: Shape estimation from shading, defocus, and correspondence using light-field angular coherence. Tao
G: Line assisted light field triangulation and stereo matching.Z. Yu,

Real-world Results

B: Globally consistent depth labeling of 4D lightfields. S. Wanner and B. Goldluecke C: Accurate depth map estimation from a lenslet light field camera.H.-G. Jeon,
D: Robust light field depth estimation for noisy scene with occlusion. W. Williem E: Occlusionaware depth estimation using light-field cameras. T.-C. Wang,.
F: Shape estimation from shading, defocus, and correspondence using light-field angular coherence. Tao G: Line assisted light field triangulation and stereo matching.Z. Yu,

Real-world Results

Center View

Center View

A: Accurate depth map estimation from a lenslet light field camera.H.-G. Jeon, B: Robust light field depth estimation for noisy scene with occlusion. W. Williem
C: Occlusionaware depth estimation using light-field cameras. T.-C. Wang,. D: Shape estimation from shading, defocus, and correspondence using light-field angular coherence. Tao
E: Line assisted light field triangulation and stereo matching.Z. Yu,

Optical Flow

Optical Flow

- Applications
  - Motion based segmentation
  - Structure from Motion(3D shape and Motion)
  - Alignment (Global motion compensation)
    - Camcorder video stabilization
    - UAV Video Analysis
  - Video Compression

FlowNet: Learning Optical Flow with Convolutional Networks

FischerPhilipp*, *Alexey Dosovitskiy, *Eddy Ilg, **Philip Häusser, **Caner Hazırbaş, **Vladimir Golkov, **Patrick van der Smagt, **Daniel Cremers, and *Thomas Brox

*University of Freiburg, **Technical University of Munich

Problem Statement

- Given the pair of images, predict the optical flow field using CNNs.

Details of the Proposed Approach

- The network is trained end-to-end.
- Contracting part and expanding part

- Contracting part extracts a rich feature representation

- Expanding Part
  - Refinement
  - Produces the high resolution flow

FlowNetSimple

- Process two stacked input images jointly.

FlowNetCorr

- First process the images separately, and then correlate their features at different locations and process further.

Refinement

- Makes use of up-convolutional layers and features from contracting part

Flying Chairs

- Existing datasets are too small to train the CNN
- To provide enough training images, they generate a synthetic dataset called Flying Chairs.
- Collected images from Flicker and available rendered set of 3D chair model.
- Add images of multiple Chair to the background

Figure 7. Examples of optical flow prediction on the Sintel dataset. In each row left to right: overlaid image pair, ground truth flow and 3 predictions: EpicFlow, FlowNetS and FlowNetC. Endpoint error is shown for every frame. Note that even though the EPE of FlowNets is usually worse than that of EpicFlow, the networks often better preserve fine details.

FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks

https://youtu.be/JSzUdVBmQP4

SELF-SUPERVISION

Strong Supervision

Weak Supervision

Self-Supervision

EXAMPLES OF SELF-SUPERVISION

I Context

II Color

III Motion

plane approaching zr
avro regional jet rj

IV Ambient Noise & Noisy Labels

Unsupervised CNN for Single View Depth Estimation: Geometry to the Rescue

Unsupervised CNN for Single View Depth Estimation: Geometry to the Rescue

Back to Basics: Unsupervised Learning of Optical Flow via Brightness Constancy and Motion Smoothness