Transfer Learning

CSE 598 Introduction to Deep Learning

Materials from https://ruder.io/transfer-learning/,
https://ruder.io/state-of-transfer-learning-in-nlp/,
https://docs.google.com/presentation/d/1fIhGikFPnb7G5kr58OvYC3GN4io7MznnM0aAgadvJfc/edit#slide=id.g5888218f39_177_4 (Tutorial on transfer learning at NAACL 2019)

Transfer Learning

- Traditional ML: learn one task at a time from scratch
  - sufficient data?
  - what if we change the domain?

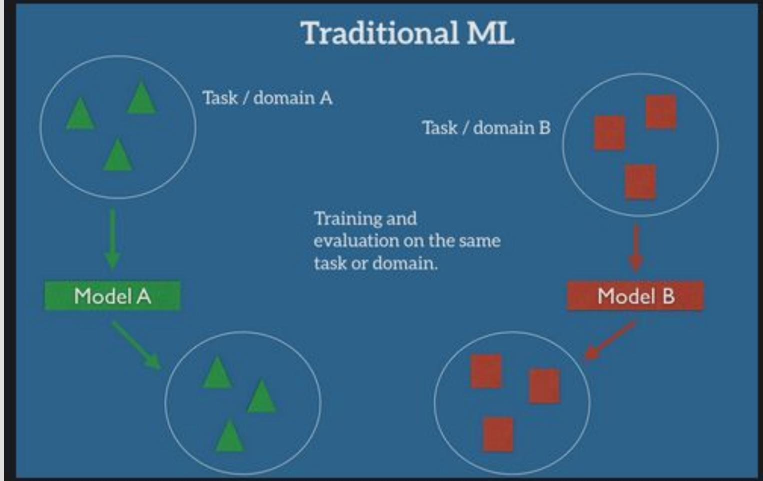


Figure 1: The traditional supervised learning setup in ML

Transfer Learning

- Transfer Learning: use whatever you have learned from a task to learn another task
    - The tasks have to be related to a certain degree
        - learn languages
        - pedestrian detection in daytime and nighttime

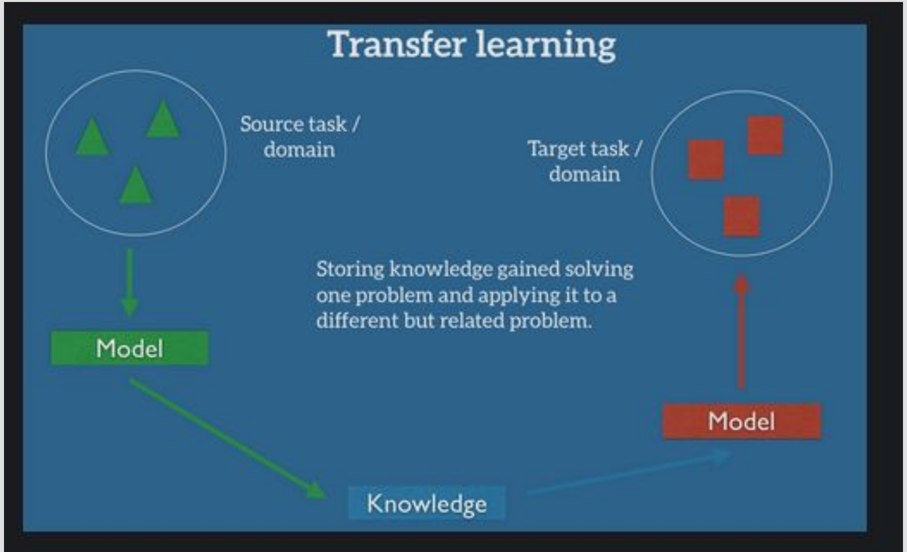
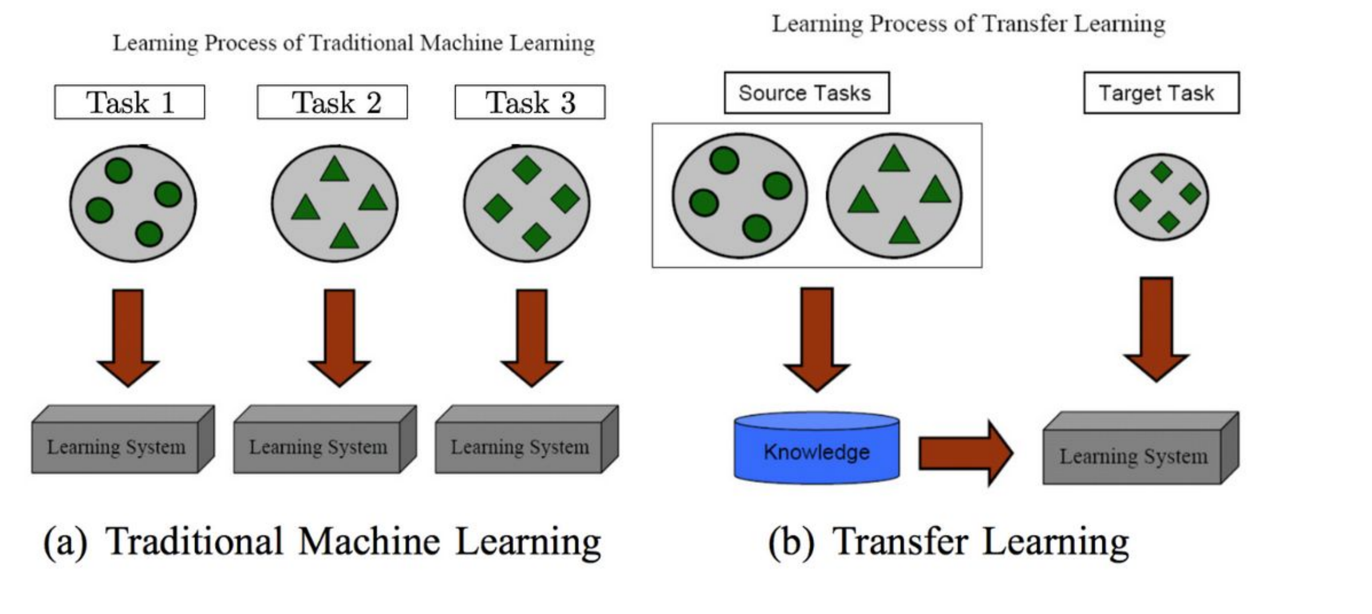


Figure 2: The transfer learning setup

Transfer Learning

- Transfer Learning: use whatever you have learned from a task to learn another task
  - The tasks have to be related to a certain degree
    - learn languages
    - pedestrian detection in daytime and nighttime



Transfer Learning

- Transfer Learning: use whatever you have learned from a task to learn another one
  - The tasks have to be related to a certain degree
    - learn languages
    - pedestrian detection in daytime and nighttime

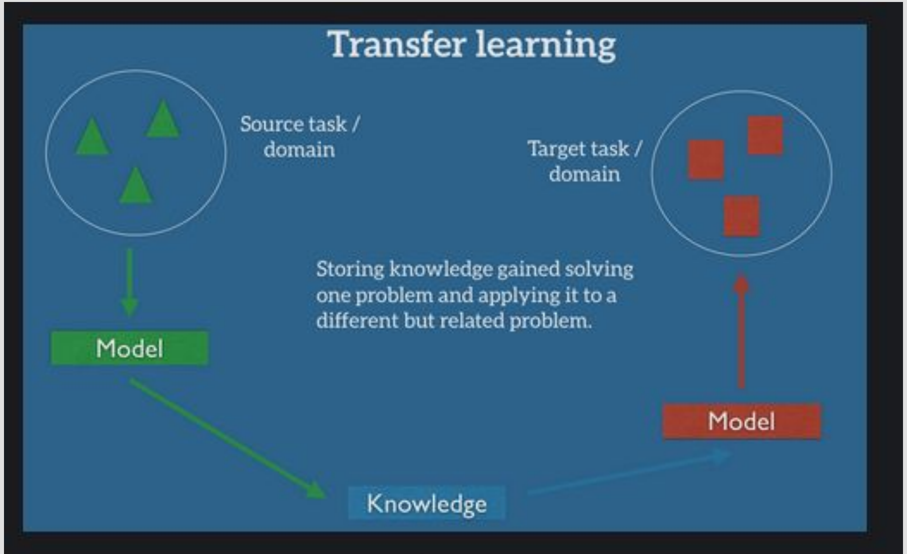


Figure 2: The transfer learning setup

Transfer Learning

Applications:

- Simulations are cheaper than interacting with the real world

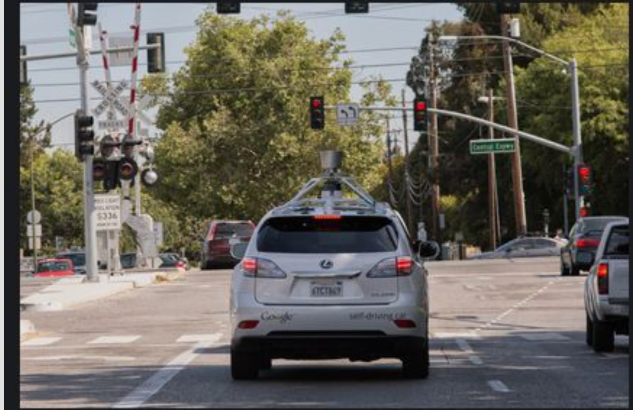


Figure 6: A Google self-driving car (source: Google Research blog)

Figure 7: Udacity's self-driving car simulator (source: TechCrunch)

Transfer Learning

Applications:

- Simulations are cheaper than interacting with the real world

Figure 8: Robot and simulation images (Rusu et al., 2016)

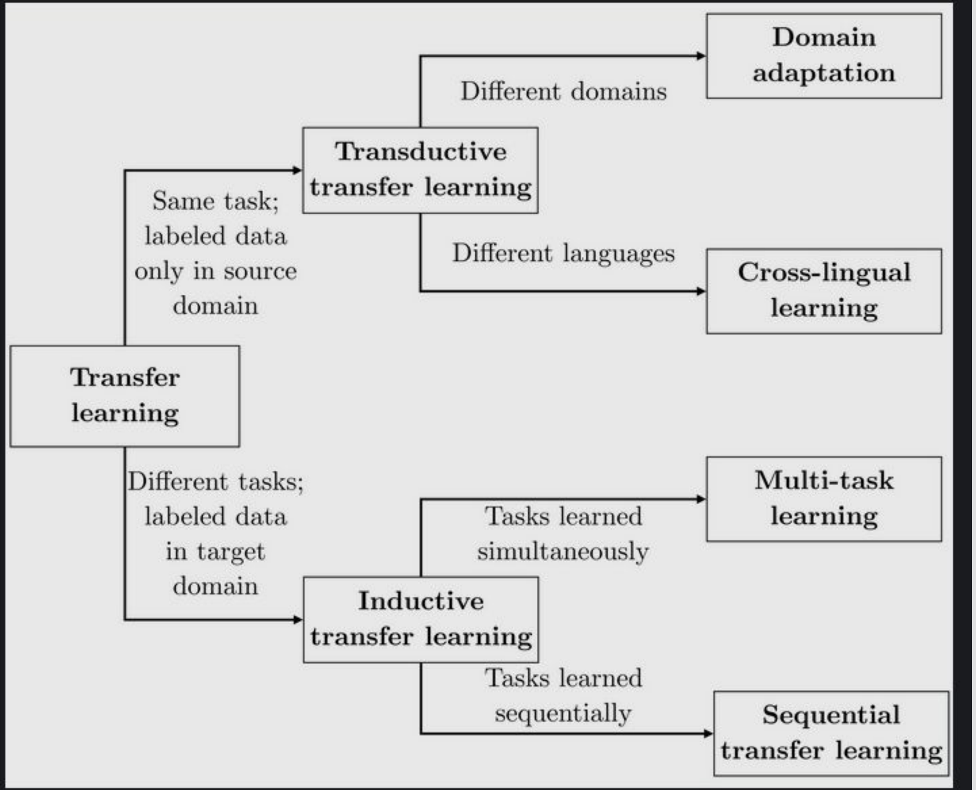
Transfer Learning

Applications:

- Adapt to new domains

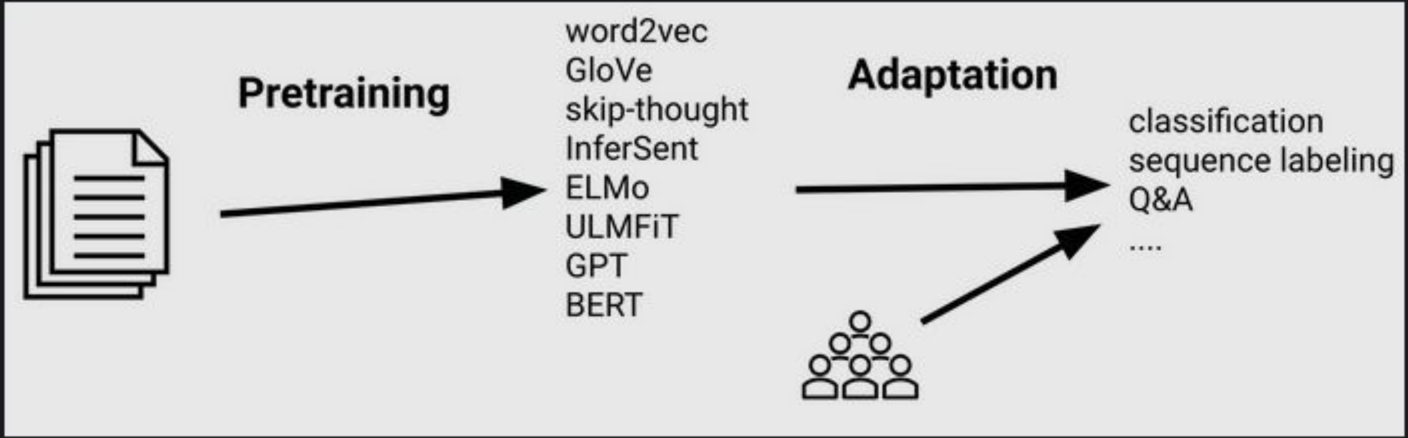
Figure 10: Different visual domains (Sun et al., 2016)

Transfer Learning



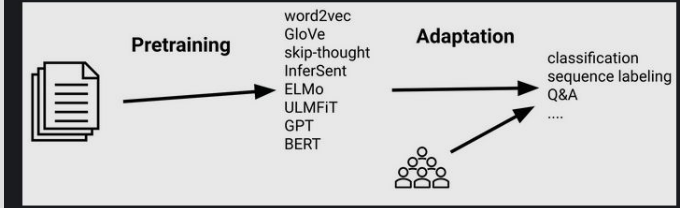
A taxonomy for transfer learning in NLP (Ruder, 2019).

Transfer Learning



The general procedure of sequential transfer learning.

Transfer Learning



The general procedure of sequential transfer learning.

- Can pretraining on related tasks help?
  - **You want to solve Task B**
  - Can pretraining with Task A1 help? What about Task A2? (regardless of the architecture)
    - It depends, but it is likely to help if Tasks Ax are related
  - Think of it like initializing weights with Task Ax
    - the standard initialization (random or whatever) has no knowledge of anything
      - the weights are not useful to solve any (known) task
    - it is probably better to start with weights that are useful for something
      - (as long as the something is related to what you care about)
      - learning sequentially

Transfer Learning - Inductive, Sequential

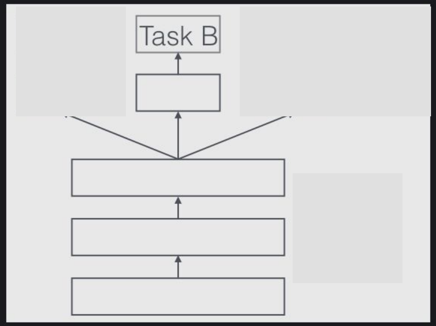


Figure 1: Hard parameter sharing for multi-task learning in deep neural networks

1. Training with Task B and nothing else

Transfer Learning - Inductive, Sequential

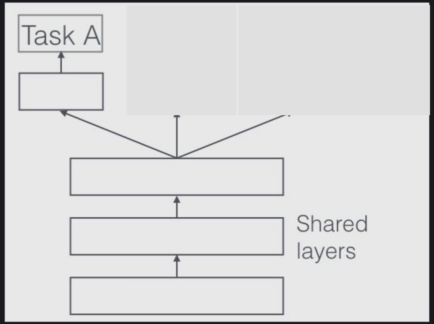


Figure 1: Hard parameter sharing for multi-task learning in deep neural networks

2a. Pretraining with A

Transfer Learning - Inductive, Sequential

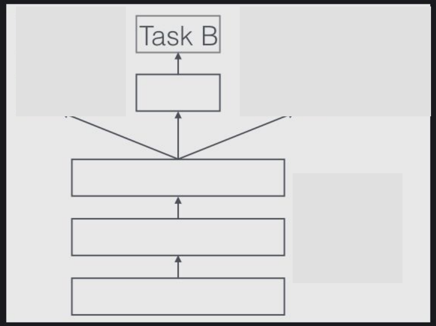


Figure 1: Hard parameter sharing for multi-task learning in deep neural networks

2b. and then fine-tuning with Task B (the task you care about)

Transfer Learning - Inductive, Sequential

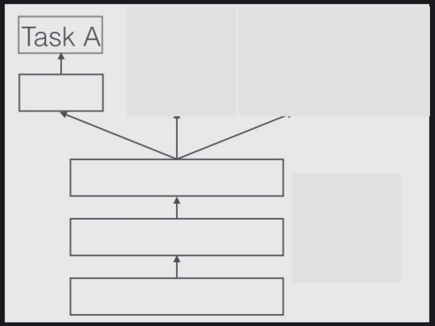


Figure 1: Hard parameter sharing for multi-task learning in deep neural networks

3a. You can also pretrain with several tasks (Task A, Task C, etc.)

Transfer Learning - Inductive, Sequential

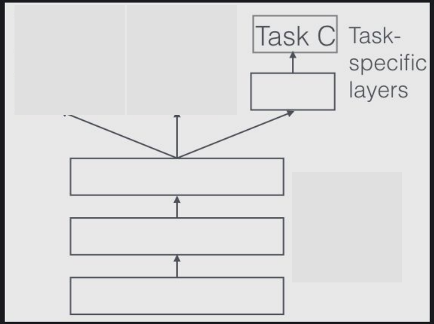


Figure 1: Hard parameter sharing for multi-task learning in deep neural networks

3a. You can also pretrain with several tasks (Task A, Task C, etc.)

Transfer Learning - Inductive, Sequential

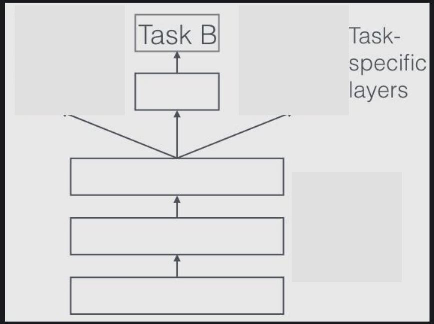


Figure 1: Hard parameter sharing for multi-task learning in deep neural networks

3b. and then fine-tuning with Task B (the task you care about)

Transfer Learning - Inductive, Multi-Task

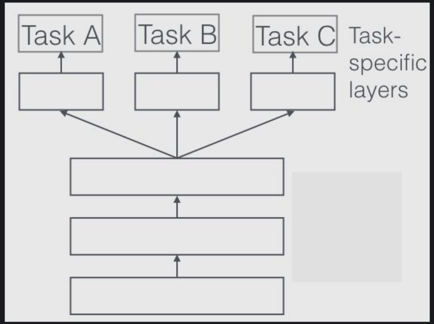


Figure 1: Hard parameter sharing for multi-task learning in deep neural networks

Transfer Learning - Inductive, Multi-Task

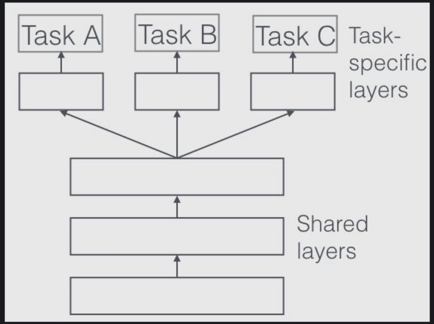


Figure 1: Hard parameter sharing for multi-task learning in deep neural networks

- And you can also train with all the tasks at once (the one you care about and auxiliary tasks)
  - Multi-Task Learning
- You will have
  - three losses (one per task)
  - each loss is backpropagated through the task-specific layer and the shared layers
  - the weights of the shared layers are updated according to the three losses! (one backpropagation step through the shared layers)

Transfer Learning

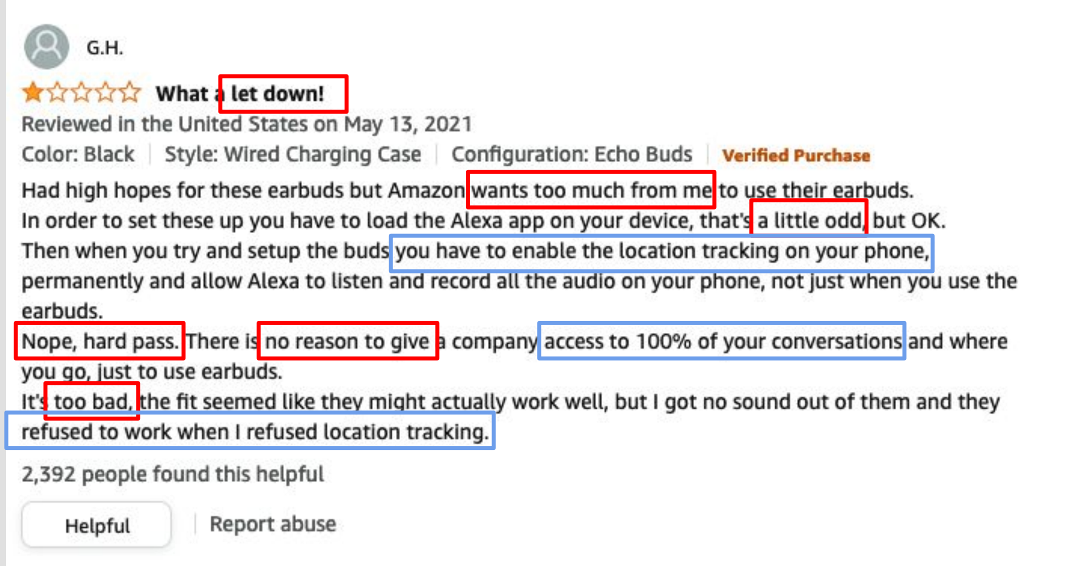
- How do choose auxiliary tasks?
    - Well, first of all you need a dataset that is annotated
    - Don't choose randomly: it it more likely to work if you use your human intuition

- Same task in a different domain
    - For example, sentiment in movie reviews vs. sentiment in Twitter
    - Relatively easy, although label definitions may change

- Different tasks are more interesting

Transfer Learning

- Same task in a different domain
  - For example, sentiment in movie reviews vs. sentiment in Tw
  - Relatively easy, although label definitions may change



G.H.

What a let down!

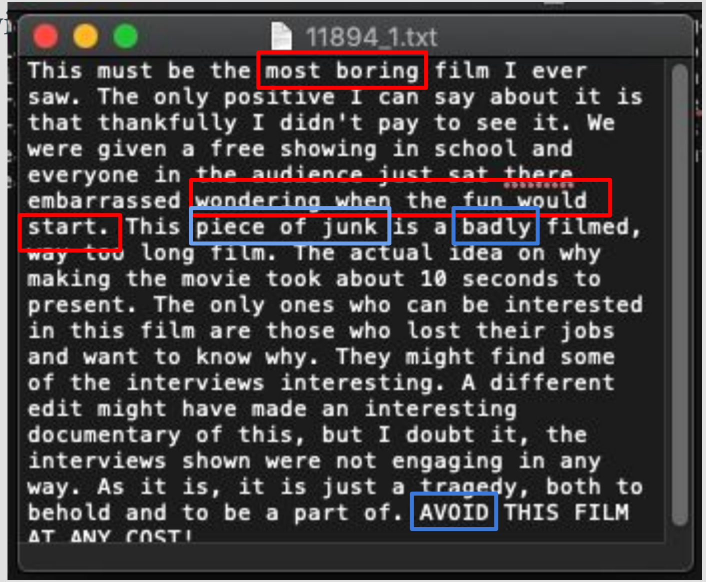
Reviewed in the United States on May 13, 2021

Color: Black | Style: Wired Charging Case | Configuration: Echo Buds | **Verified Purchase**

Had high hopes for these earbuds but Amazon wants too much from me to use their earbuds.
In order to set these up you have to load the Alexa app on your device, that's a little odd, but OK.
Then when you try and setup the buds you have to enable the location tracking on your phone, permanently and allow Alexa to listen and record all the audio on your phone, not just when you use the earbuds.
Nope, hard pass. There is no reason to give a company access to 100% of your conversations and where you go, just to use earbuds.
It's too bad, the fit seemed like they might actually work well, but I got no sound out of them and they refused to work when I refused location tracking.

2,392 people found this helpful

Helpful | Report abuse



11894_1.txt

This must be the most boring film I ever saw. The only positive I can say about it is that thankfully I didn't pay to see it. We were given a free showing in school and everyone in the audience just sat there embarrassed wondering when the fun would start. This piece of junk is a badly filmed, way too long film. The actual idea on why making the movie took about 10 seconds to present. The only ones who can be interested in this film are those who lost their jobs and want to know why. They might find some of the interviews interesting. A different edit might have made an interesting documentary of this, but I doubt it, the interviews shown were not engaging in any way. As it is, it is just a tragedy, both to behold and to be a part of. AVOID THIS FILM AT ANY COST!

Answering Yes-No Questions

"I'd rather just go to bed": Understanding Indirect Answers

Annie Louis
Google Research, UK
annielouis@google.com

Dan Roth[*]
University of Pennsylvania
danroth@seas.upenn.edu

Filip Radlinski
Google Research, UK
filiprad@google.com

S1: Talking to a friend about food preferences. Q: "Do you like pizza?" A: "I like it when the toppings are meat, not vegetable."
S2: Talking to a friend about music preferences. Q: "Do you like guitars?" A: "I practice playing each weekend."
S3: Talking to a friend about weekend activities. Q: "Are you available this Sunday evening?" A: "What did you have in mind?"
S4: Talking to a friend about book preferences. Q: "Are you a fan of comic books?" A: "I read an Archie every time I have lunch."
S5. Your friend is visiting from out of town. Q: "Would you like to go out for dinner?" A: "I could go for some Mexican."
S6. Two colleagues leaving work on a Friday. Q: "Long week?" A: "I've had worse weeks."
S7. You friend is planning to buy a flat in New York. Q: "Does the flat's price fit your long-term budget?" A: "Well, if it doesn't I will definitely refinance my mortgage."
S8. Your friend is thinking of switching jobs. Q: "Do you have to travel far?" A: "My commute is about 10 minutes."
S9. Two childhood neighbours uexpectedly run into each other at a cafe. Q: "Are you going to the high school reunion in June?" A: "I forgot all about that."
S10. Meeting your new neighbour for the first time. Q: "Did you move from near-by?" A: "I am from Canada."

Table 2: Examples of questions and answers in our 10 dialogue scenarios.

Answering Yes-No Questions

-

Yes Q: Do you have any pets? A: My cat just turned one year old.	**Probably yes / sometimes yes** Q: Do you like mysteries? A: I have a few that I like.	**Yes, subject to some conditions** Q: Do you enjoy drum solos? A: When someone's a master.
No Q: Do you have a house? A: We are in a 9th floor apartment.	**Probably no** Q: Are you interested in fishing this weekend? A: It's supposed to rain.	**In the middle** Q: Did you find this week good? A: It was the same as always.

Answering Yes-No Questions

-

Model	Accuracy Dev.	Accuracy Test
		Bas
Majority class	50.2	49.3
MNLI	28.4	28.9
BOOLQ	64.2	**62.7**
		BERT finetu
BERT-YN (Question only)	56.4	56.0
BERT-YN (Answer only)	83.0	**81.7**
		BERT finet
BERT-YN	88.4	<u>87.8</u>
BERT-MNLI-YN	89.6	**88.2**
BERT-DIS-YN	88.0	87.4
BERT-BOOLQ-YN	87.7	87.1

Answering Yes-No Questions

- MNLI

Examples

Premise	Label	Hypothesis
Fiction		
The Old One always comforted Ca'daan, except today.	*neutral*	Ca'daan knew the Old One very well.
Letters		
Your gift is appreciated by each and every student who will benefit from your generosity.	*neutral*	Hundreds of students will benefit from your generosity.
Telephone Speech		
yes now you know if if everybody like in August when everybody's on vacation or something we can dress a little more casual or	*contradiction*	August is a black out month for vacations in the company.
9/11 Report		
At the other end of Pennsylvania Avenue, people began to line up for a White House tour.	*entailment*	People formed a line at the end of Pennsylvania Avenue.

Answering Yes-No Questions

Examples

Premise	Label	Hypothesis
Fiction		
The Old One always comforted Ca'daan, except today.	*neutral*	Ca'daan knew the Old One very well.
Letters		
Your gift is appreciated by each and every student who will benefit from your generosity.	*neutral*	Hundreds of students will benefit from your generosity.
Telephone Speech		
yes now you know if if everybody like in August when everybody's on vacation or something we can dress a little more casual or	*contradiction*	August is a black out month for vacations in the company.
9/11 Report		
At the other end of Pennsylvania Avenue, people began to line up for a White House tour.	*entailment*	People formed a line at the end of Pennsylvania Avenue.

- MNLI

- Using NNLI to answering yes-no questions
  - Mapping inputs:
    - Q: Do you like Italian food
    - A: I love Tuscan food
    - Premise: I like Italian food (declarative version of the question)
    - Hypotheses: I love Tuscan food (entailment)
  - Mapping labels:
    - entailment -> YES
    - contradiction -> NO
    - neutral -> In the middle

Answering Yes-No Questions

- BoolQ

BoolQ

Passage: *Barq's – Barq's is an American soft drink. Its brand of root beer is notable for having caffeine. Barq's, created by Edward Barq and bottled since the turn of the 20th century, is owned by the Barq family but bottled by the Coca-Cola Company. It was known as Barq's Famous Olde Tyme Root Beer until 2012.*

Question: *is barq's root beer a pepsi product* **Answer:** `No`

Answering Yes-No Questions

- YN - Circa, their corpus

Yes Q: Do you have any pets? A: My cat just turned one year old.	**Probably yes / sometimes yes** Q: Do you like mysteries? A: I have a few that I like.	**Yes, subject to some conditions** Q: Do you enjoy drum solos? A: When someone's a master.
No Q: Do you have a house? A: We are in a 9th floor apartment.	**Probably no** Q: Are you interested in fishing this weekend? A: It's supposed to rain.	**In the middle** Q: Did you find this week good? A: It was the same as always.

Answering Yes-No Questions

Allen Nie[*1] **Erin D. Bennett**[*2] **Noah D. Goodman**[1,2]
[1]Department of Computer Science [2]Department of Psychology
Stanford University
anie@cs.stanford.edu {erindb,ngoodman}@stanford.edu

- DIS - discourse connective prediction task.
  - The predict the best marker between declarative version of the question and the indirect answer
    - only *because, but, if, when, and*

S1	marker	S2
Her eyes flew up to his face.	and	Suddenly she realized why he looked so different.
The concept is simple.	but	The execution will be incredibly dangerous.
You used to feel pride.	because	You defended innocent people.
Ill tell you about it.	if	You give me your number.
Belter was still hard at work.	when	Drade and barney strolled in.
We plugged bulky headsets into the dashboard.	so	We could hear each other when we spoke into the microphones.
It was mere minutes or hours.	before	He finally fell into unconsciousness.
And then the cloudy darkness lifted.	though	The lifeboat did not slow down.

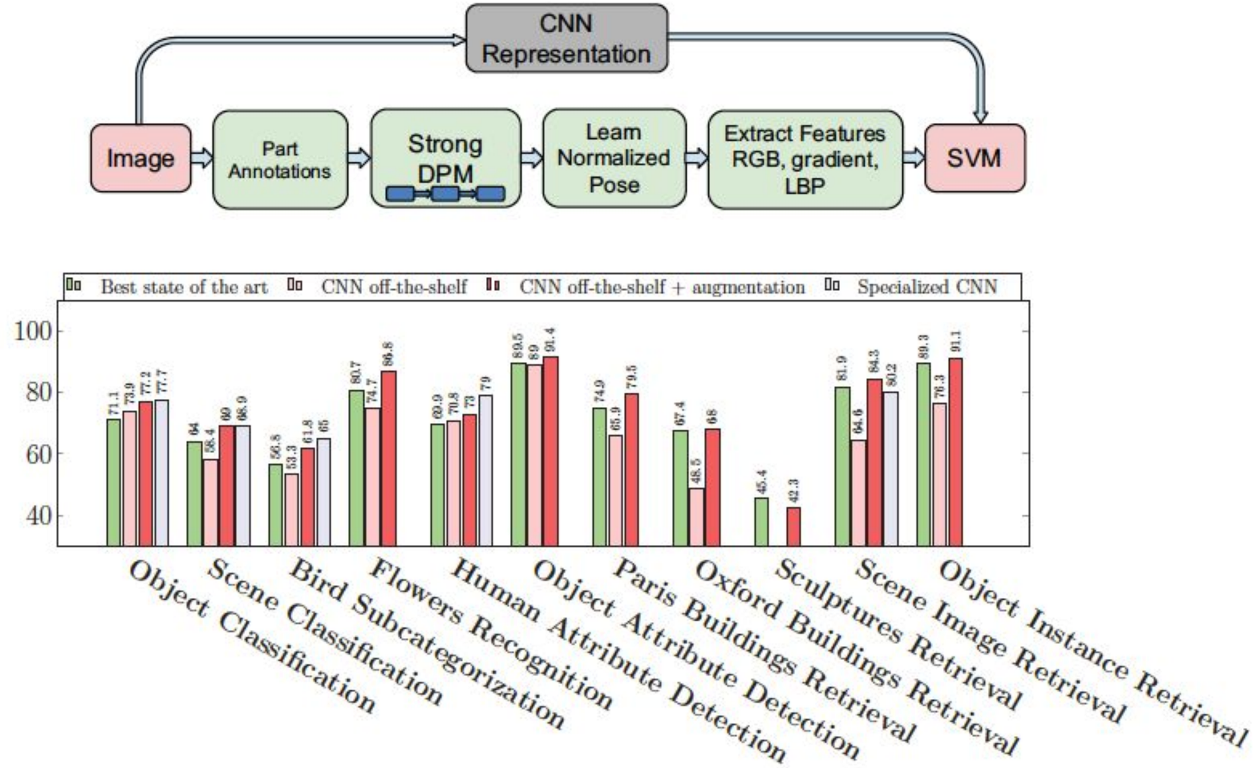
Table 3: **Example pairs** from our Books 8 dataset.

CNN Features off-the-shelf: an Astounding Baseline for Recognition

Razavian, A., Hossein Azizpour, J. Sullivan and S. Carlsson. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (2014). https://openaccess.thecvf.com/content_cvpr_workshops_2014/W15/papers/Razavian_CNN_Features_Off-the-Shelf_2014_CVPR_paper.pdf

- What kind of representation do they propose for images?
- Where do the features come from?

CNN Features off-the-shelf: an Astounding Baseline for Recognition



Multi-Task Video Captioning

Pasunuru, Ramakanth and Mohit Bansal. Multi-Task Video Captioning with Video and Entailment Generation. ACL (2017). https://www.aclweb.org/anthology/P17-1117/

- Describe the task of video captioning in terms of input and output.
- What baselines does the paper present?
- What tasks do they use in their many-to-many multi-task learning model (Figure 4)?
- What do you think of the human evaluation?

Multi-Task Video Captioning

-

Ground truth: A person is mixing powdered ingradients with water.
A woman is mixing flour and water in a bowl.
Our model: A person is mixing ingredients in a bowl.

Figure 1: A video captioning example from the YouTube2Text dataset, with the ground truth captions and our many-to-many multi-task model's predicted caption.

Multi-Task Video Captioning

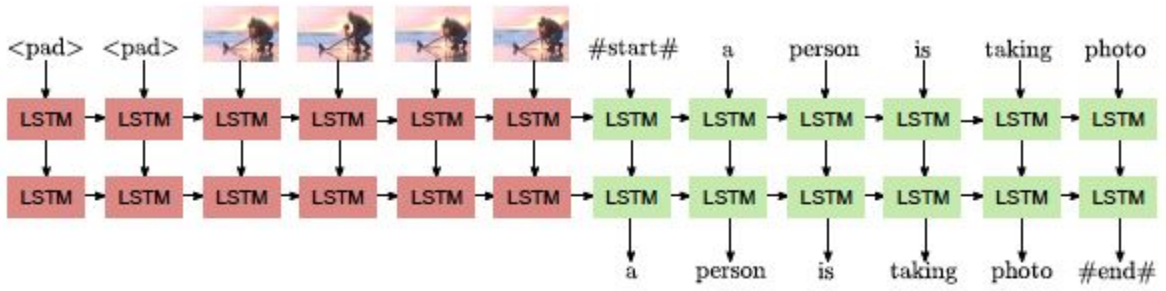


Figure 2: Baseline sequence-to-sequence model for video captioning: standard encoder-decoder LSTM-RNN model.

Multi-Task Video Captioning

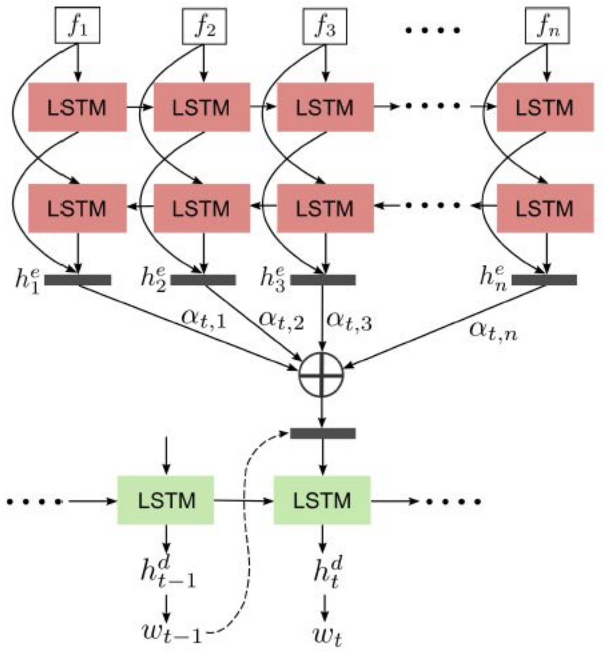


Figure 3: Attention-based sequence-to-sequence baseline model for video captioning (similar models also used for video prediction and entailment generation).

Multi-Task Video Captioning

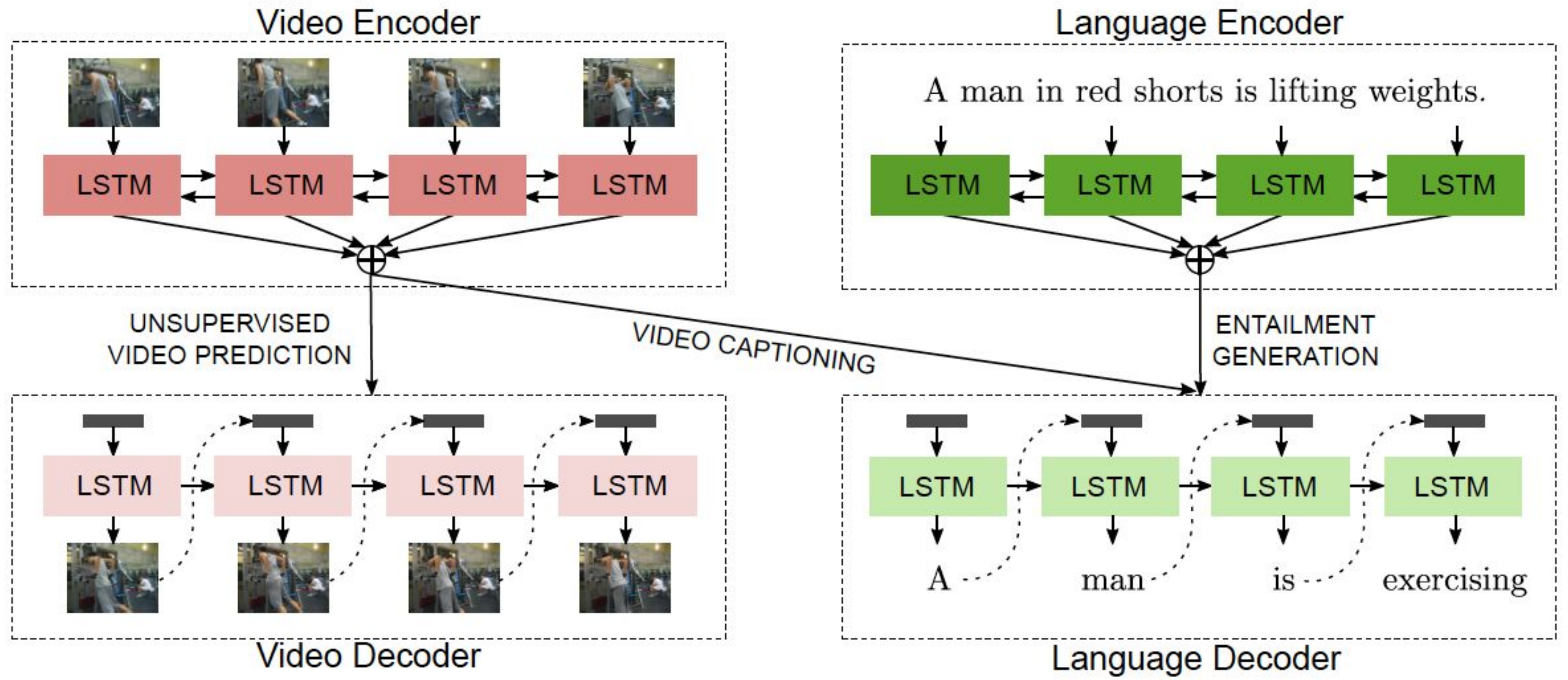


Figure 4: Our many-to-many multi-task learning model to share encoders and decoders of the video captioning, unsupervised video prediction, and entailment generation tasks.