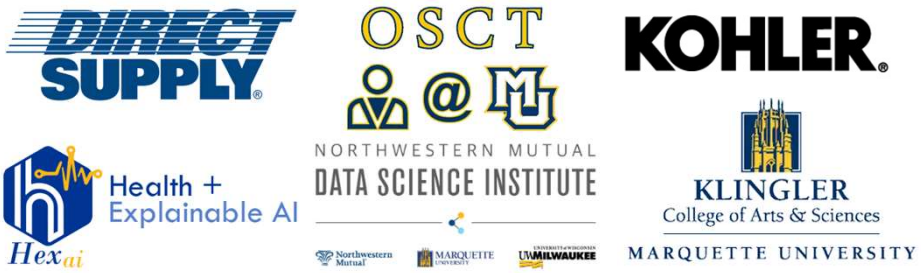
Transfer Learning
in
Computer Vision
and
Natural Language Processing



1

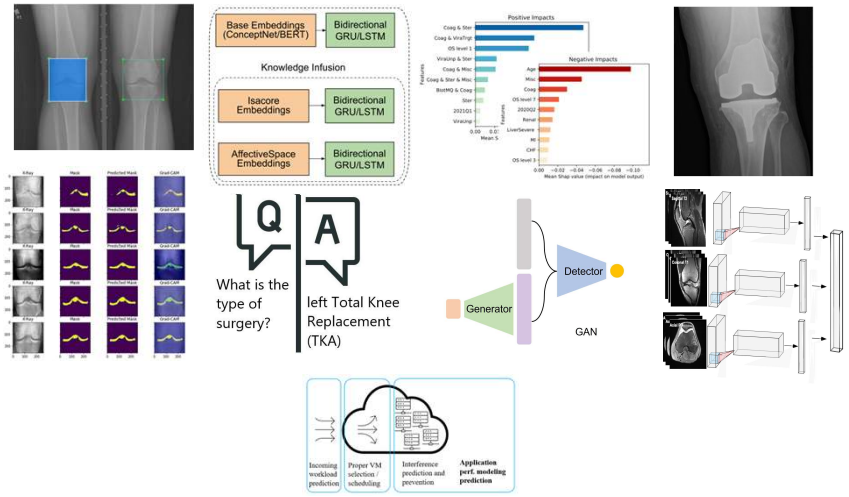
Brief Introduction

Hamidreza Moradi
Assistant Professor
Department of Data Science
Director, XDI Lab
University of Mississippi Medical Center

2

2

Brief Introduction

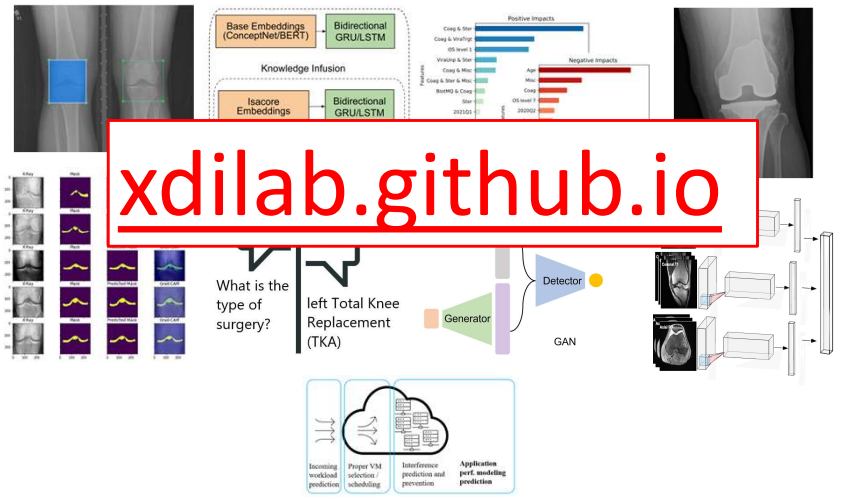


MISSISSIPPI CENTER FOR CLINICAL AND TRANSLATIONAL RESEARCH · National COVID Cohort Collaborative · UMMC · IDeA CTRnet

3

3

Brief Introduction



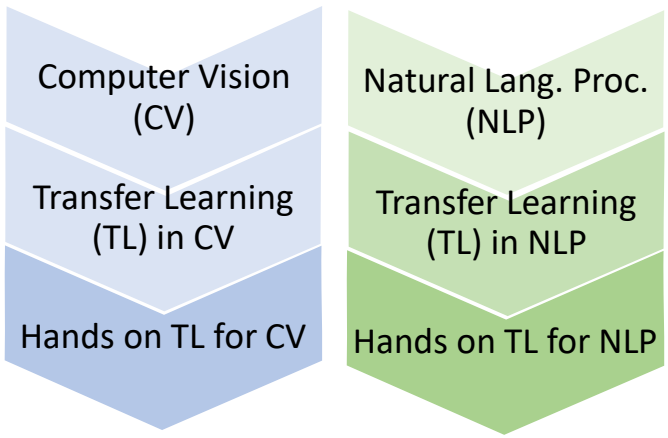
xdilab.github.io

MISSISSIPPI CENTER FOR CLINICAL AND TRANSLATIONAL RESEARCH · National COVID Cohort Collaborative · UMMC · IDeA CTRnet

4

4

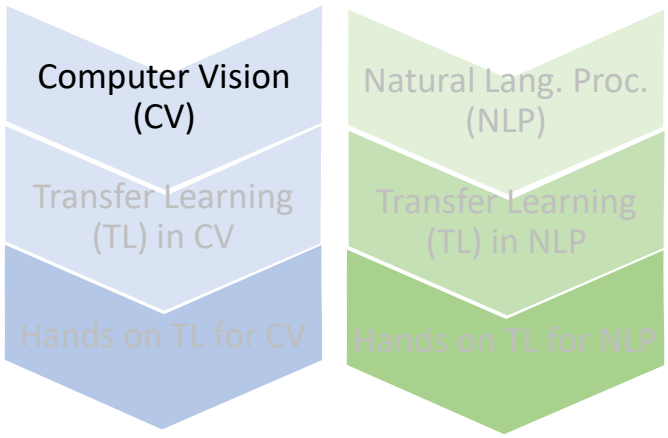
Outline



5

5

Outline

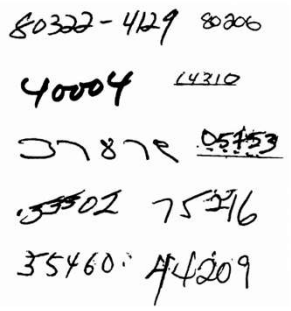


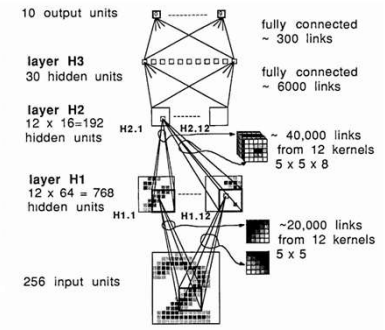
6

6

The Beginning

- Yann LeCun (1989): Recognize zip codes for US postal service





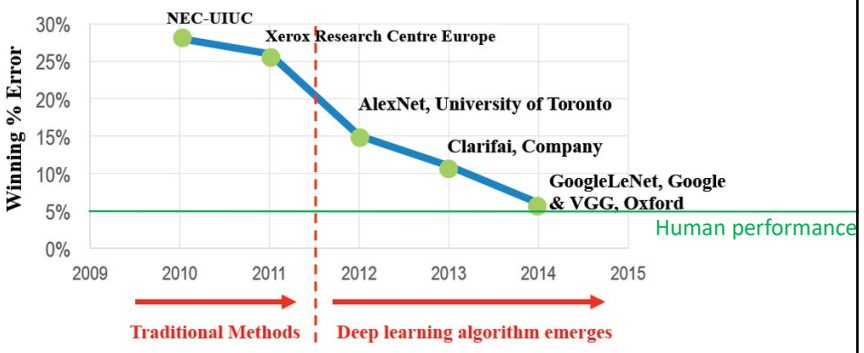
- Geoffrey Hinton (2006): same with accuracy>98%

7

7

The Excitement

- AlexNet (2012): trained on 200 GB of ImageNet Data

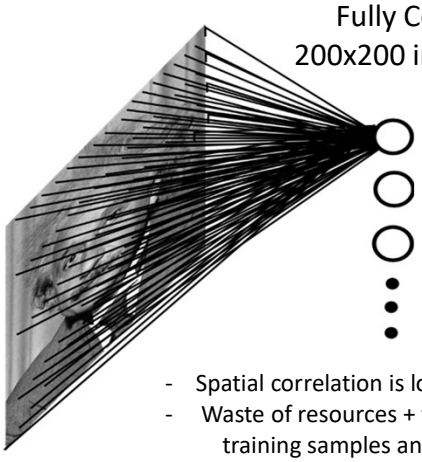


2018 Shawahna et al. 8

8

Fully Connected Layer for Visual Data

Fully Connected layer from
200x200 images to multiple hidden units



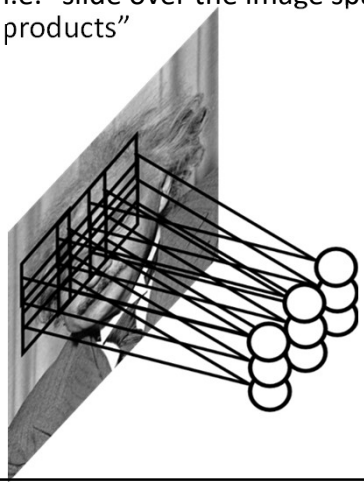
- Spatial correlation is local
- Waste of resources + we have not enough training samples anyway..

9

9

Convolution Layer

- Convolve filters over the image
  - i.e. “slide over the image spatially, computing dot products”

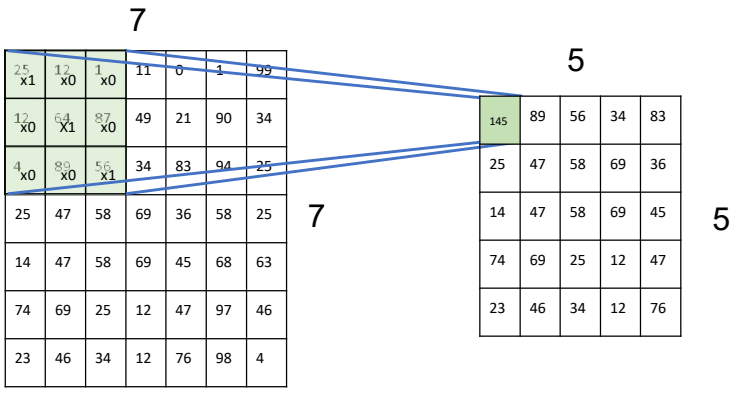


10

10

Convolution Layer (cont.)

- Convolve filters over the image
  - i.e. "slide over the image spatially, computing dot products"

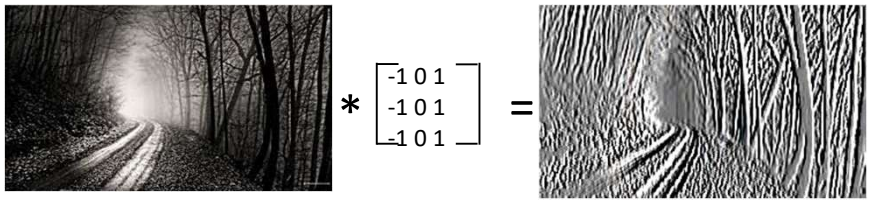


11

11

Convolution Layer (cont.)

- Convolve filters over the image
  - i.e. "slide over the image spatially, computing dot products and then sum the values"

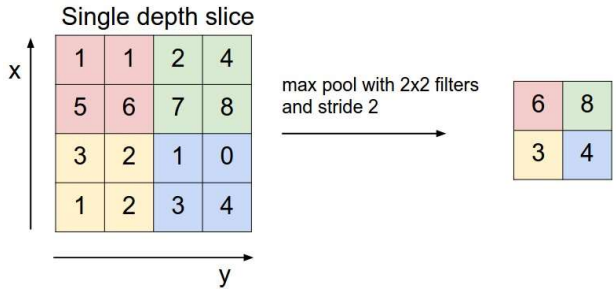


12

Pooling

Max pooling
minimum pooling
average pooling

- Progressively reduce the spatial size
  - Reduce the amount of parameters
  - Reduce computation in the network
  - Also control overfitting



13

13

The Landscape

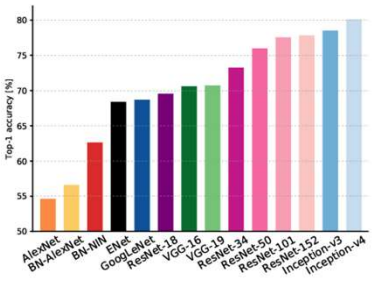


Figure 1: **Top1 *vs.* network.** Single-crop top-1 validation accuracies for top scoring single-model architectures. We introduce with this chart our choice of

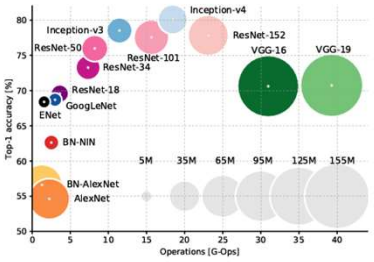


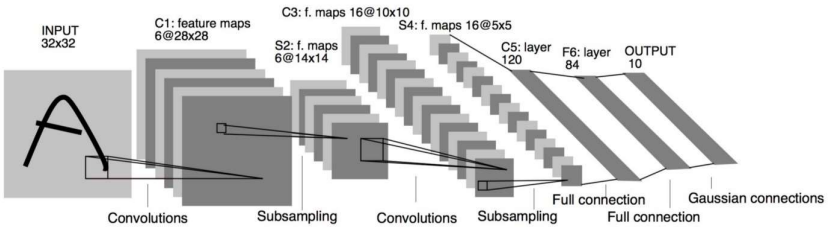
Figure 2: **Top1 *vs.* operations, size $\propto$ parameters.** Top-1 one-crop accuracy versus amount of operations required for a single forward pass. The size of the

Paper: An analysis of DNN models, Canziani & Culurciello et al. 14

14

LeNet5

- Year 1994, first convolutional neural networks
  - after many previous successful iterations since 1988
- Pioneering work by Yann LeCun
- With the insight that image features are distributed across the entire image



15

Parameters: 60,000

15

1998 to 2010 incubation Period

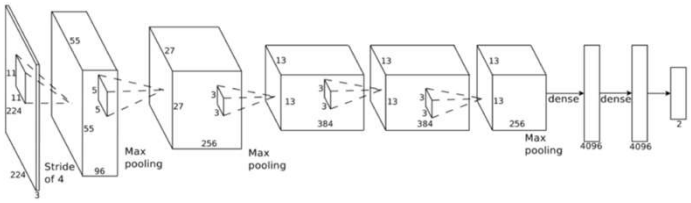
In 2010 fist GPU implementation on NVIDIA GTX 280

16

16

AlexNet

- Year 2012, by Alex Krizhevsky For ImageNet Competition
- Deeper and wider version of the LeNet
- contribution of this work:
  - Use of rectified linear units (ReLU) as non-linearities
  - Use of dropout technique during training to avoid overfitting



Parameters: 60,000

17

17

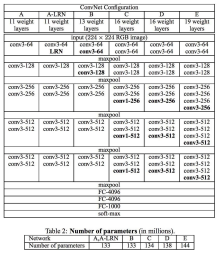
VGG

- Use of 3×3 filters in each convolutional layers
  - LeNet: large convolutions to capture similar features
  - AlexNet: 9×9 or 11×11 filters
  - Multiple 3×3 convolution in sequence can emulate the effect of 5×5 and 7×7.
- Filters started to become smaller
  - Close to the infamous 1×1 convolutions that LeNet wanted to avoid

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers

Table 2: **Number of parameters** (in millions).

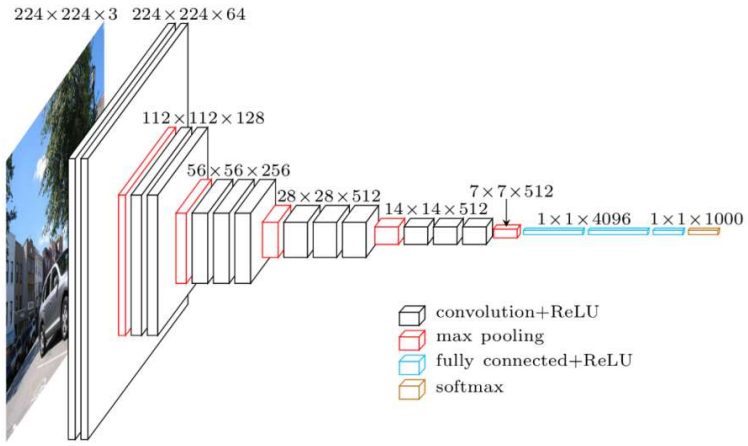
Network	A,A-LRN	B	C	D	E
Number of parameters	133	133	134	138	144



18

18

VGG-16



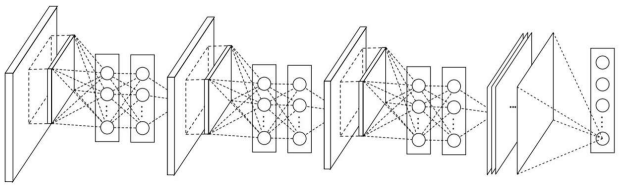
Parameters: 138 million

19

19

Network-in-network (NiN)

- Using 1x1 convolutions to provide more combinational power
- MLP layers after each convolution
  - Better combine features before another layer
  - Different from using raw pixels
  - Spatially combine features across features maps after convolution
- Average pooling layer as part of the last classifier



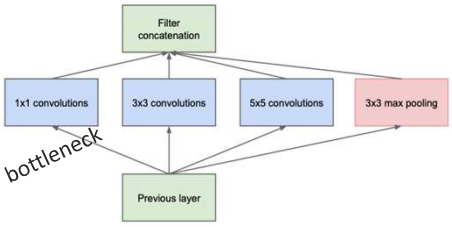
20

20

Inception (GoogLeNet)

- Year 2014, deep learning models used in categorizing the content of images
- By Christian Szegedy to Reduce the computational
- They came up with the Inception module:

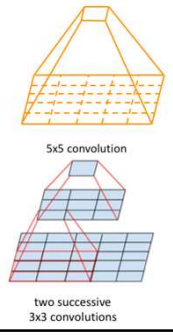
- parallel combination of 1×1, 3×3, 5x5
- Use of 1×1 convolutional blocks to reduce the number of features

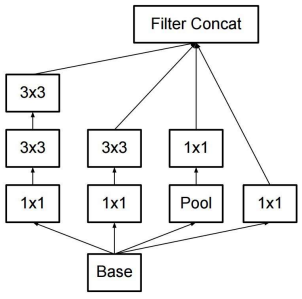


Naïve Implementation

Inception V2 & V3

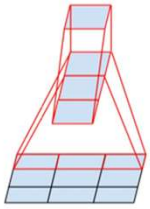
- Feb. 2015
  - Batch-normalized Inception (V2)
- Dec. 2015, (V3)
  - Use only 3x3 convolution
  - Filter of 5x5 and 7x7 can be decomposed

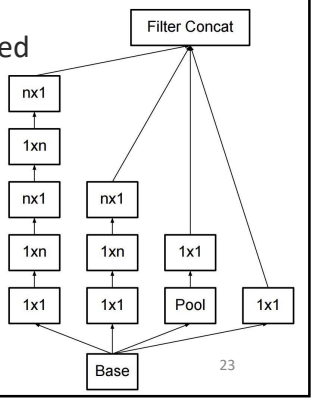




Inception V2 & V3

- Feb. 2015
  - Batch-normalized Inception (V2)
- Dec. 2015, (V3)
  - Use only 3x3 convolution
  - Filter of 5x5 and 7x7 can be decomposed





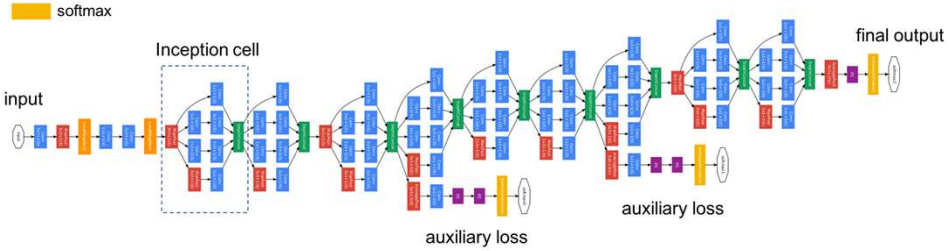
23

23

Inception V2 & V3

convolution
max pooling
channel concatenation
channel-wise normalization
fully-connected layer
softmax

Papers
- Going deeper with convolutions
- Rethinking the Inception Architecture for Computer Vision



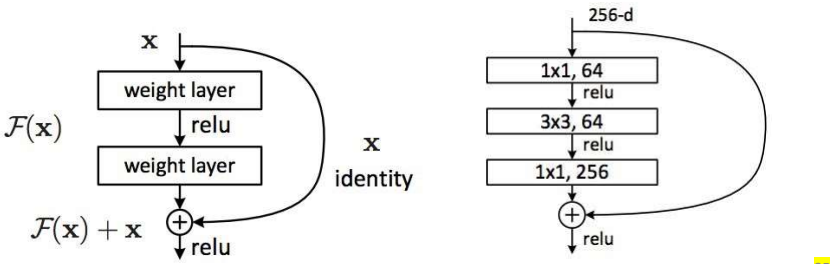
Parameters: 5 million (V1) and 23 million (V3)

24

24

ResNet

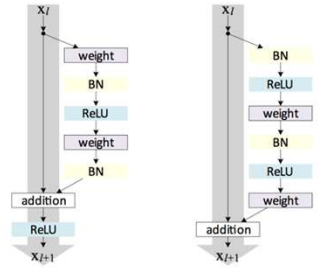
- Dec. 2015, same time as Inception v3
  - Feed the output of two successive convolutional layer AND also bypass the input to the next layers
  - Bypassing after 2 layers is a key intuition
    - As a **small classifier**, or a Network-In-Network
  - First time that a network of > hundred trained



25

25

ResNet



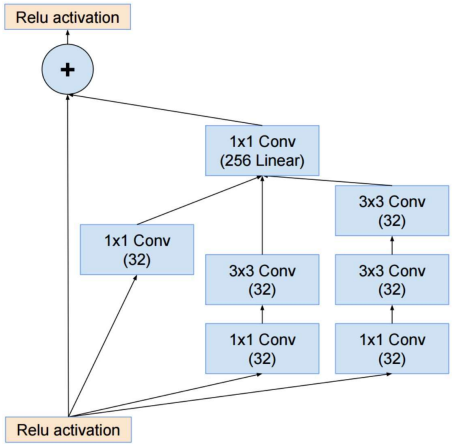
original residual unit proposed revision

Parameters: 25 million (ResNet 50)

26

Inception V4

- Combined the Inception module with the ResNet module

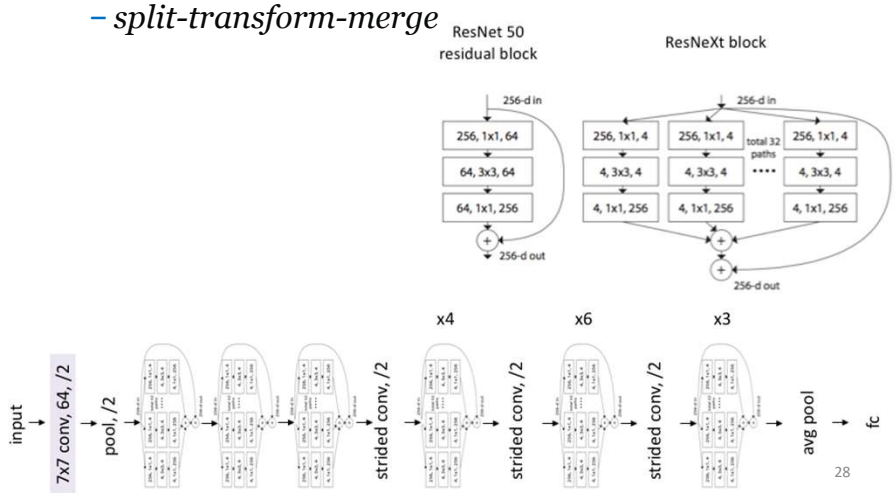


27

27

ResNeXt

- Extension of the deep residual network
  - *split-transform-merge*

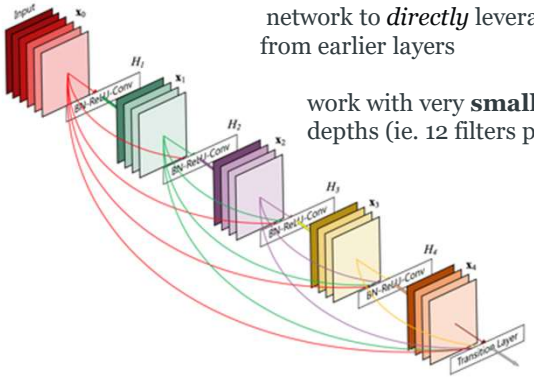


28

28

DenseNet

- Useful to reference feature maps from earlier in the network
  - Each layer's feature map is concatenated to the input of *every successive layer*



network to *directly* leverage the features from earlier layers

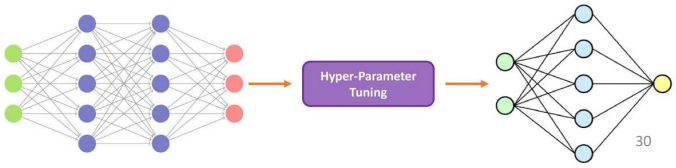
work with very **small output channel** depths (ie. 12 filters per layer)

29

29

Hyper-parameters for CNN

- Kernel/Filter Size
  - In general we use filters with odd sizes
- Padding
  - Output size shrinks to (n-f)/s+1 where 'n' is input dimensions 'f' is filter size and 's' is stride length (Output with padding?)
- Stride
- Number of Channels
- Pooling-layer Parameters



30

30

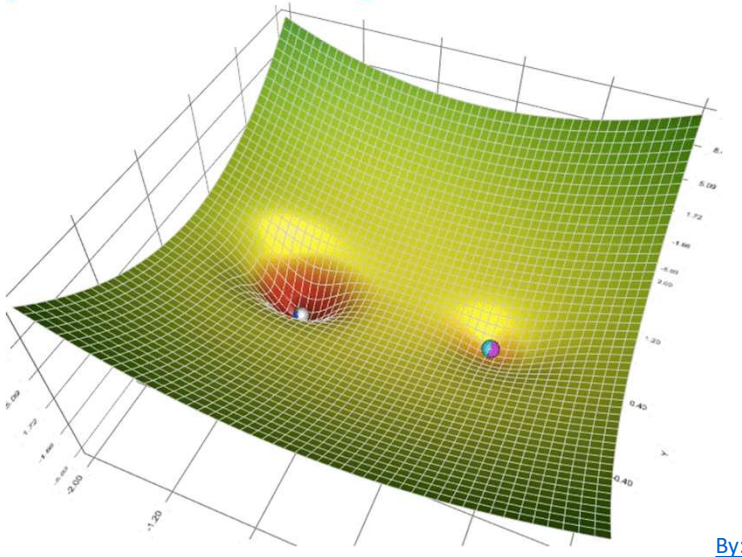
Optimization Algorithms

- **Momentum:**
  - In each step, in addition to the regular gradient, also adds on the movement from the previous step*** Decay Factor**
- **Adagrad :**
  - keeps track of the sum of gradient *squared* and uses that to adapt the gradient (SLOW)
- **RMSprop:**
  - Fixes Adagrad issue by adding a decay factor
- **Adam:**
  - takes the best of both worlds of Momentum and RMSProp

31

31

Optimization Algorithms



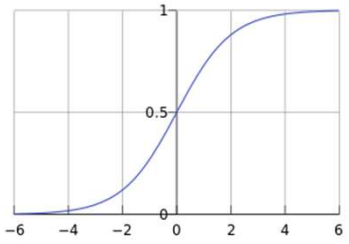
By: Lili Jiang

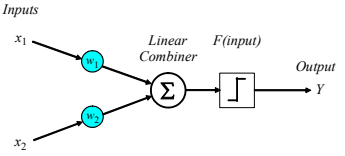
gradient descent (cyan), momentum (magenta), AdaGrad (white), RMSProp (green), Adam (blue)

32

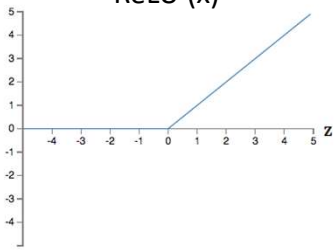
Nonlinearity with activation functions

Sigmoid / Logistic Function

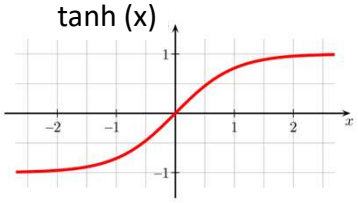




ReLU (x)



tanh (x)



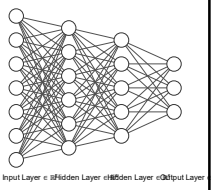
Leaky ReLU, Parametric ReLU, ELU

List of activation functions: https://www.tensorflow.org/api_docs/python/tf/nn

33

33

Weight Initialization



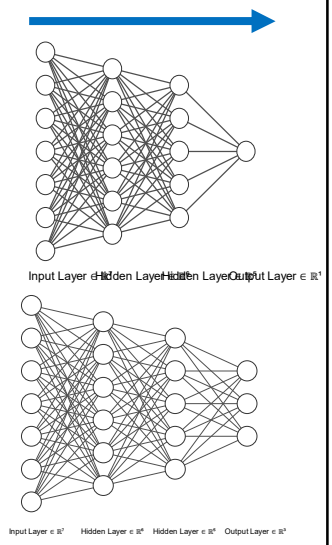
- Initialization matter
- Historically: Random Weight initialization
  - Heuristic Gaussian or uniform distribution
  - Small random values in the range [-0.3, 0.3]
  - Small random values in the range [0, 1]
  - Small random values in the range [-1, 1]
- "Glorot" or "Xavier" initialization (2010 by Xavier Glorot)
  - Based on the assumption that the activations are linear
  - Good Sigmoid or TanH activation function, invalid for ReLU
  - Uniform distribution in range of [-(1/sqrt(n)), 1/sqrt(n)]
- "he" initialization (2015 by Kaiming He)
  - Gaussian distribution, mean of 0.0 and Std of sqrt(2/n)

34

34

Loss Function

- Regression:
  - Same loss as Linear Regression
  - Quadratic loss (i.e. mean squared error)
- Classification:
  - Use the same objective as Logistic Regression
  - Cross-entropy (i.e. negative log likelihood)
  - Binary: $-(y\log(p) + (1-y)\log(1-p))$
  - Multi class: $-\sum_{c=1}^{M} y_{o,c}\log(p_{o,c})$

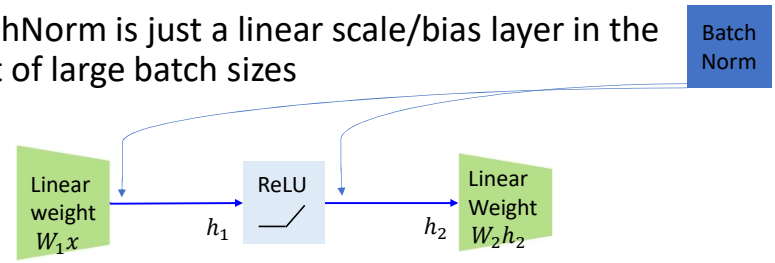


35

35

Batch normalization [Ioffe and Szegedy, 2015]

- BatchNorm is just a linear scale/bias layer in the limit of large batch sizes



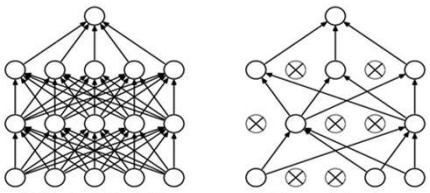
- Improves gradient flow through the network
- Allows higher learning rates
- Reduces the strong dependence on initialization
- Reduces need for regularization

36

36

Dropout Regularization [Srivastava et al., 2014]

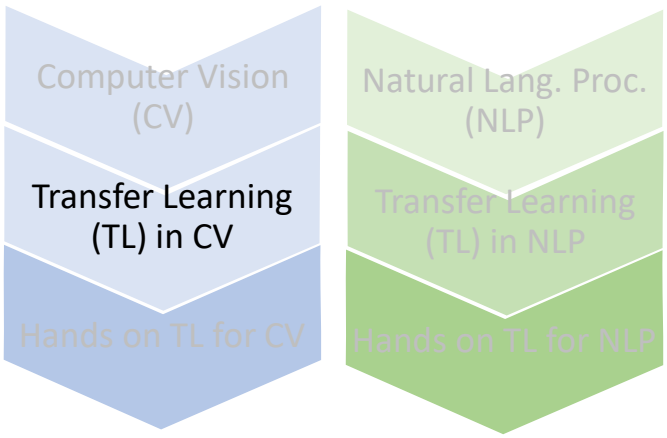
- Dropout: form of regularization, useful for NNs
- Works by randomly "dropping out" units in a network for a single gradient step
- The more you drop out, the stronger the regularization
  - 0.0 = no dropout regularization
  - 1.0 = drop everything out! learns nothing


(a) Standard Neural Net (b) After applying dropout.

37

37

Outline



38

38

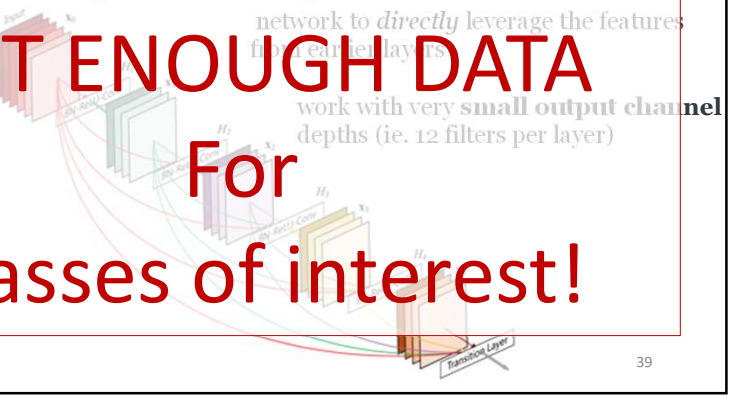
DenseNet

Trained for a certain Task!

NOT ENOUGH DATA

For

Classes of interest!

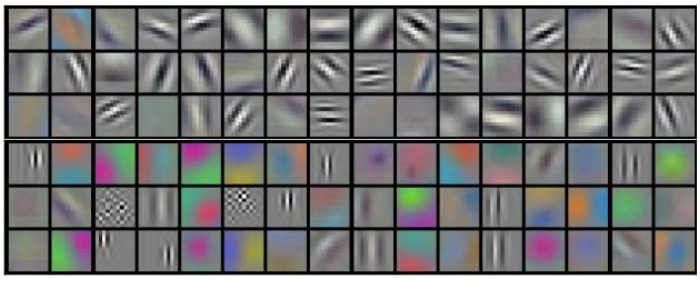


39

39

Features Extracted by CNN

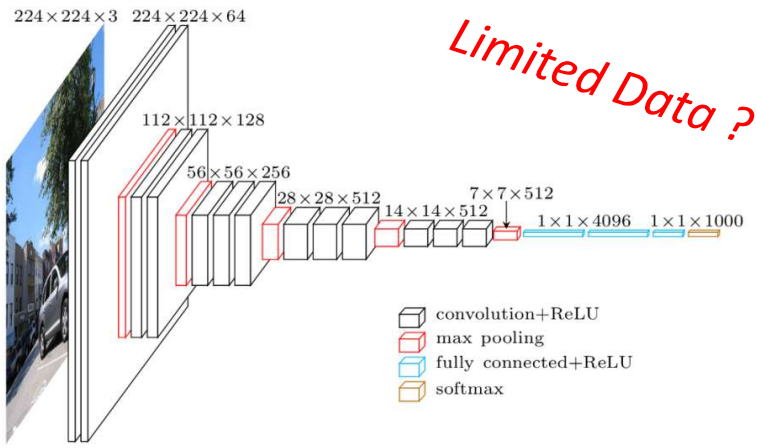
- Example filters learned by Krizhevsky et al. Each of the 96 filters shown here is of size [11x11x3],



40

40

Fine Tune Fully connected Layers

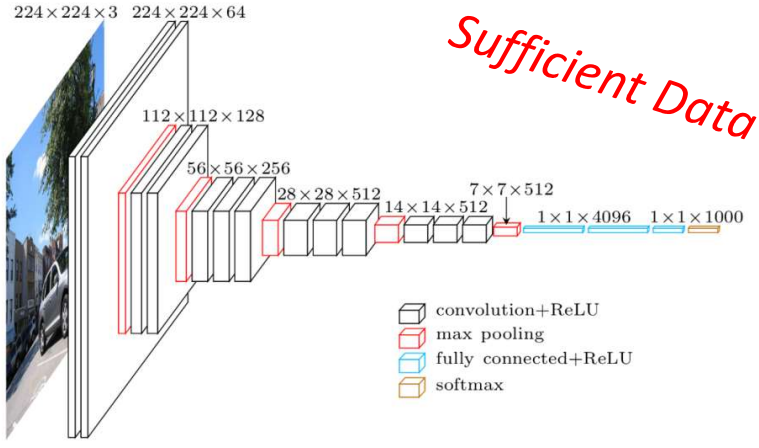


Limited Data ?

41

41

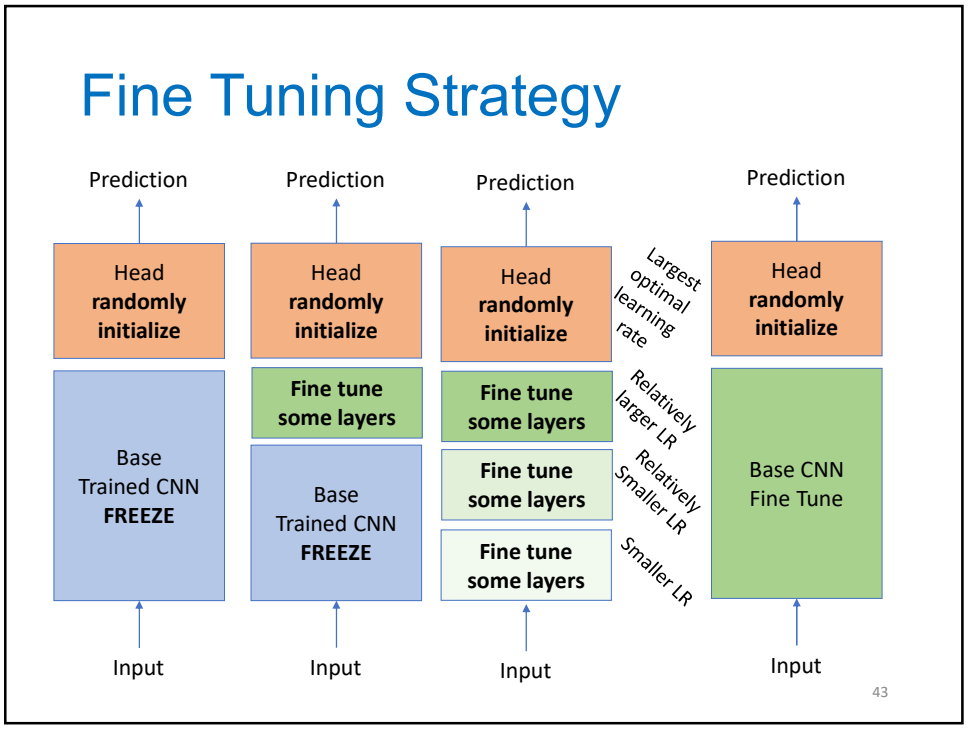
Fine Tune Whole Model



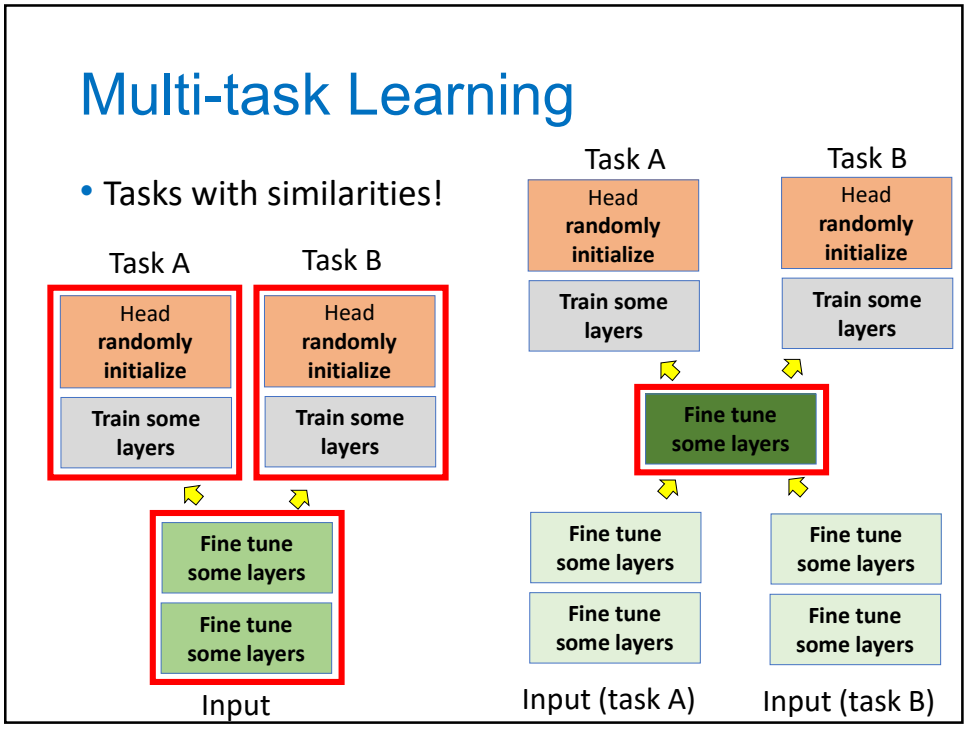
Sufficient Data ?

42

42


Fine Tuning Strategy
Prediction
Prediction
Prediction
Prediction
Head randomly initialize
Head randomly initialize
Head randomly initialize
Head randomly initialize
Largest optimal learning rate
Base Trained CNN FREEZE
Fine tune some layers
Base Trained CNN FREEZE
Fine tune some layers
Relatively larger LR
Fine tune some layers
Relatively Smaller LR
Fine tune some layers
Smaller LR
Base CNN Fine Tune
Input
Input
Input
Input
43

43

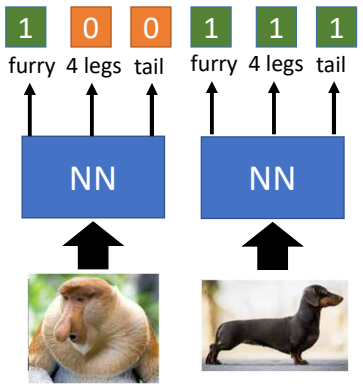

Multi-task Learning
• Tasks with similarities!
Task A
Task B
Head randomly initialize
Head randomly initialize
Train some layers
Train some layers
Fine tune some layers
Fine tune some layers
Input
Task A
Task B
Head randomly initialize
Head randomly initialize
Train some layers
Train some layers
Fine tune some layers
Fine tune some layers
Fine tune some layers
Fine tune some layers
Fine tune some layers
Input (task A)
Input (task B)

44

Zero-shot Learning

- Representing each class by its attributes

Trained model



Provides One to one mapping

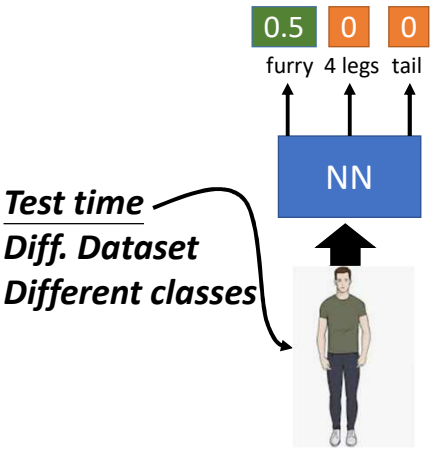
attributes

classes	furry	4 legs	tail	…
Dog	x	x	x	
Monkey	x	-	-	
…				

46

Zero-shot Learning

- Representing each class by its attributs



Class with more similarities!

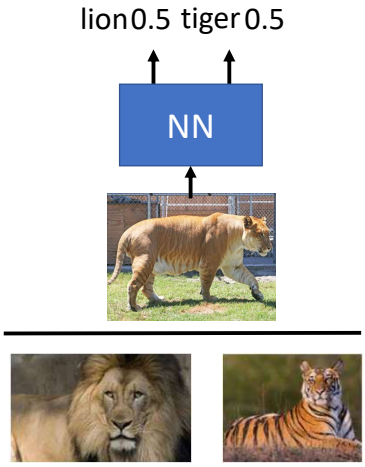
attributes

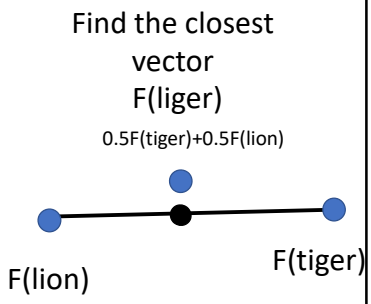
classes	furry	4 legs	tail	…
Dog	x	x	x	
Monkey	x	-	-	
…				

47

Zero-shot Learning

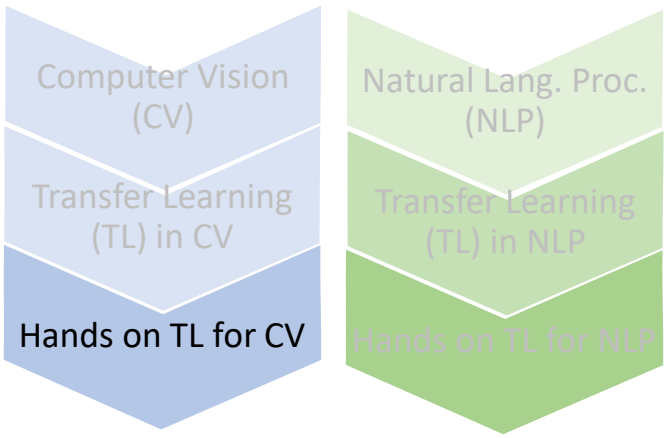
- Combination of Embedding





48

Outline



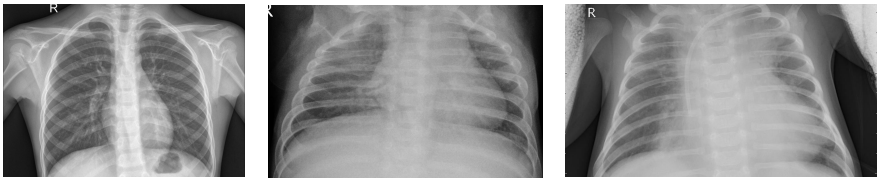
49

49

Radiograph Classification

- Pneumonia vs. Normal X-ray

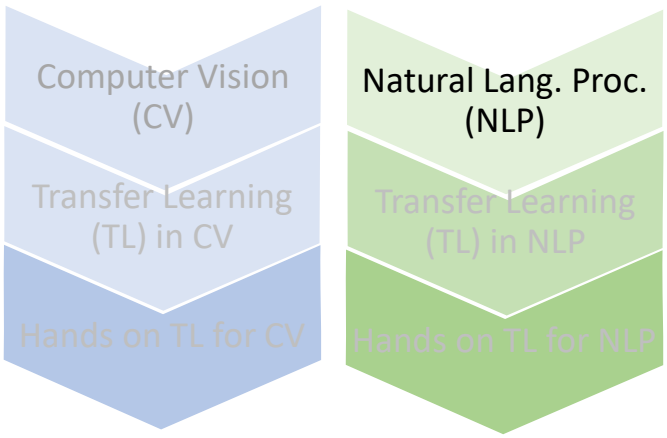
Normal | Viral | Bacterial



https://data.mendeley.com/datasets/rscbjbr9sj/2 50

50

Outline



51

51

Features

- Numerical: Good, you may need to normalize

	A	B	C	D
1	accelerations	fetal_movement	light_decelerations	fetal_health
2	0	0	0	0
3	0.006	0	0.003	1
4	0.003	0	0.003	1
5	0.003	0	0.003	1
6	0.007	0	0	1
7	0.001	0	0.009	0
8	0.001	0	0.008	0
9	0	0	0	0
10	0	0	0	0
11	0	0	0	0
12	0	0	0.001	1

Fetus Datasets

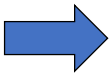
53

53

Features (cont.)

- Categorical ?

Color
Red
Green
Red
Blue
Yellow
…
Green
Missing



color-blue	color-green	color-missing	color-red	color-yellow	…
0	0	0	1	0	
0	1	0	0	0	
0	0	0	1	0	
1	0	0	0	0	
0	0	0	0	1	
0	1	0	0	0	
0	0	1	0	0	

54

54

Features (cont.)

- Using description as features?

	A	B	C
1	video_id	comment_count	description
2	2kyS6SvSYSE	15954	SHANTELL'S CHANNEL - https://www.youtube.com/shantellmartin\nCANDICE - https:/
3	1ZAPwfrtAFY	12703	One year after the presidential election, John Oliver discusses what we've learned so
4	5qpjK5DgCt4	8181	WATCH MY PREVIOUS VIDEO â–¶ \n\nSUBSCRIBE â–º https://www.youtube.com/chan
5	puqaWrEC7tY	2146	Today we find out if Link is a Nickelback amateur or a secret Nickelback devotee. GMM
6	d380meD0W0M	17518	I know it's been a while since we did this show, but we're back with what might be the
7	gHZ1Qz0KiKM	1434	Using the iPhone for the past two weeks -- here's my thoughts!\nAll my iPhone X Vide
8	39idVpFF7NQ	1970	Embattled Alabama Senate candidate Roy Moore (Mikey Day) meets with Vice Preside
9	nc99ccSXST0	3432	Ice Cream Pint Combination Lock - http://amzn.to/2ACipdI\nMini Ice Cream Sandwich
10	jr9QtXwC9vc	340	Inspired by the imagination of P.T. Barnum, The Greatest Showman is an original musi
11	TUmyygCMMGA	2368	For now, at least, we have better things to worry about.\n\n\nSubscribe to our channe
12	9wRQljFNDW8	177	New England Patriots returner Dion Lewis blasts off for an amazing kickoff return touc

55

55

What's next?

- Assume a cleaned string of input
- How can we represent a sentence?
  - Map words to a unique integer

56

56

Bag of words Featurization

Assuming we have a dictionary mapping words to a unique integer id, a bag-of-words featurization of a sentence could look like this:

Sentence:	The	cat	sat	on	the	mat
word id's:	1	12	5	3	1	14

The BoW featurization would be the vector:

Vector	2,	0,	1,	0,	1,	0,	0,	0,	0,	0,	0,	1,	0,	1
position	1		3		5							12		14

Note that the original word order is lost, replaced by the order of id's.

57

N-grams

Because word order is lost, the sentence meaning is weakened. This sentence has quite a different meaning but the same BoW vector:

Sentence:	The	mat	sat	on	the	cat
word id s:	1	14	5	3	1	12

BoW featurization:

Vector	2,	0,	1,	0,	1,	0,	0,	0,	0,	0,	0,	1,	0,	1

But word order is important, especially the order of nearby words.

N-grams capture this, by modeling **tuples of consecutive words**.

58

N-grams

Sentence: The cat sat on the mat

2-grams: the-cat, cat-sat, sat-on, on-the, the-mat

Notice how even these short n-grams "make sense" as linguistic units. For the other sentence we would have different features:

Sentence: The mat sat on the cat

2-grams: the-mat, mat-sat, sat-on, on-the, the-cat

We can go still further and construct 3-grams:

Sentence: The cat sat on the mat

3-grams: the-cat-sat, cat-sat-on, sat-on-the, on-the-mat

Which capture still more of the meaning:

Sentence: The mat sat on the cat

3-grams: the-mat-sat, mat-sat-on, sat-on-the, on-the-cat

59

N-grams Features

Typically, its advantages to use multiple n-gram features in machine learning models with text, e.g.

unigrams + bigrams (2-grams) + trigrams (3-grams).

The **unigrams** have higher counts and are able to detect influences that are weak, while **bigrams and trigrams** capture strong influences that are more specific.

e.g. "the white house" will generally have very different influences from the sum of influences of "the", "white", "house".

60

N-grams size

N-grams pose some challenges in feature set size.

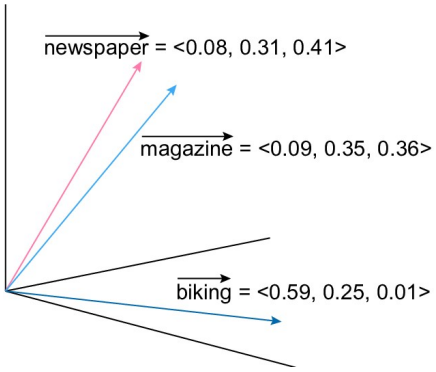
- Unigram dictionary size: 40,000
- Bigram dictionary size: 100,000
- Trigram dictionary size: 300,000

With coverage of > 80% of the features occurring in the text.

61

Word Embedding

Fancy way of saying numerical representation of words



Word2Vec (2013)
GloVe (2014)
FastText (2016)

62

Word Embedding

Fancy way of saying numerical representation of words

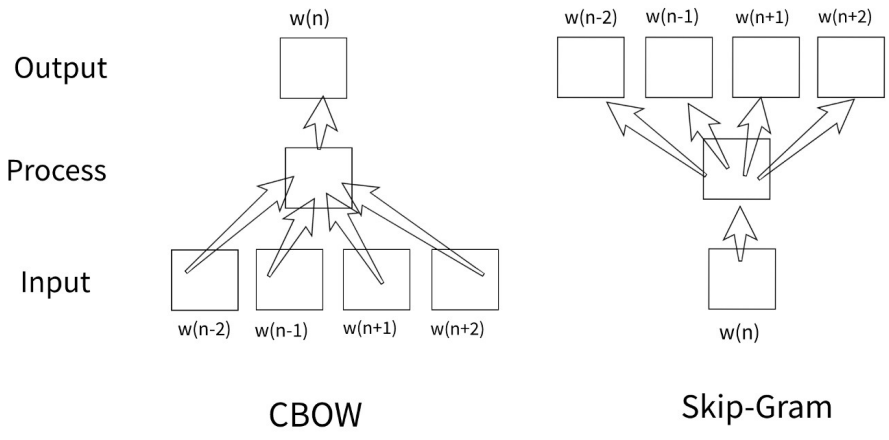
We can also **analyze the meaning of a particular word** by looking at the contexts in which it occurs.

The context is the **set of words that occur near the word** (window size), i.e. at displacements of …,-3,-2,-1,+1,+2,+3,… in each sentence where the word occurs.

2 approach:

- Skip-Gram
- Continuous Bag Of Words

63



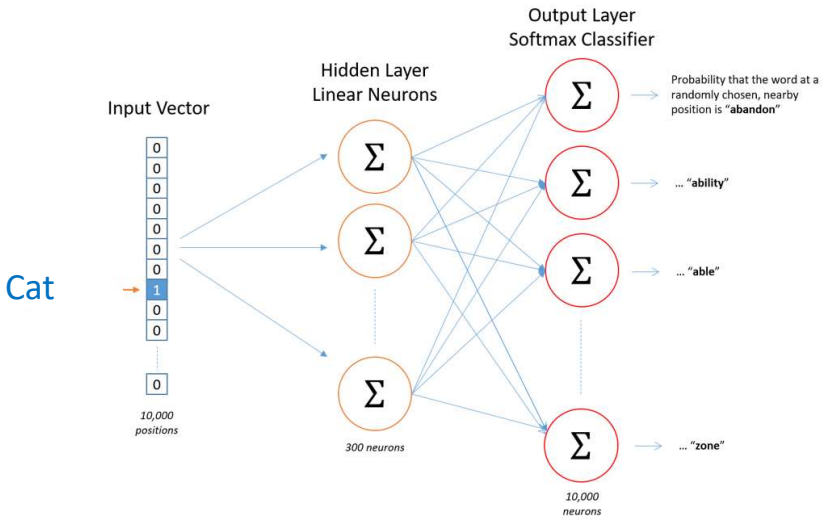
https://towardsdatascience.com/word-embedding-techniques-word2vec-and-tf-idf-explained-c5d02e34d08

https://towardsdatascience.com/nlp-101-word2vec-skip-gram-and-cbow-93512ee24314

64

64

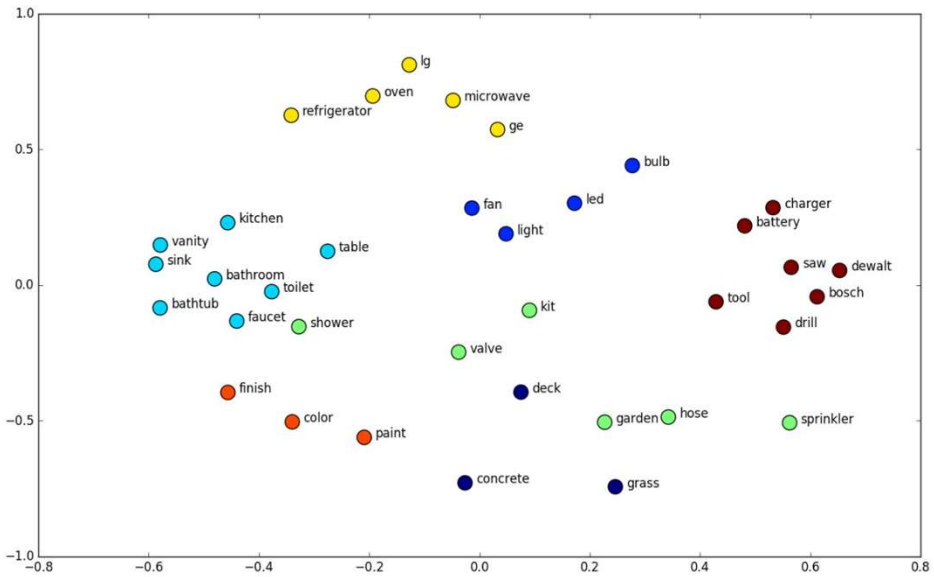
Skip-gram



http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/

65

65



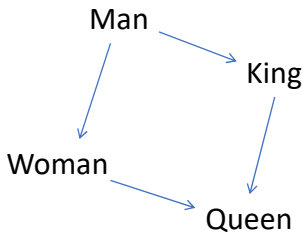
https://www.shanelynn.ie/get-busy-with-word-embeddings-introduction/

66

66

Skip-grams

Word meaning has an algebraic structure:



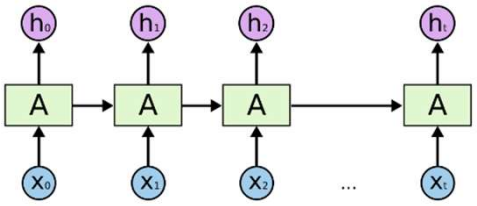
King – Man + Woman = Queen

Tomáš Mikolov et al. (2013). "Efficient Estimation of Word Representations in Vector Space"

67

Recurrent Neural Networks (RNNs)

- RNNs introduce the notion of order/time by a cycle



- Designed to process sequence data $x_1, \ldots, x_n$
- And can generate a sequence outputs $y_1, \ldots, y_m$

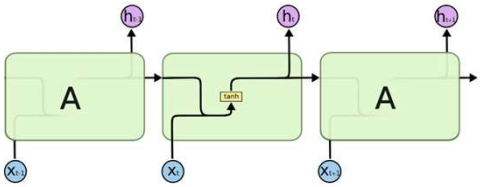
68

68

RNN: issue

The clouds are in the *sky*
I grew up in France… I speak fluent *French*

- In standard RNNs, this repeating module will have a very simple structure



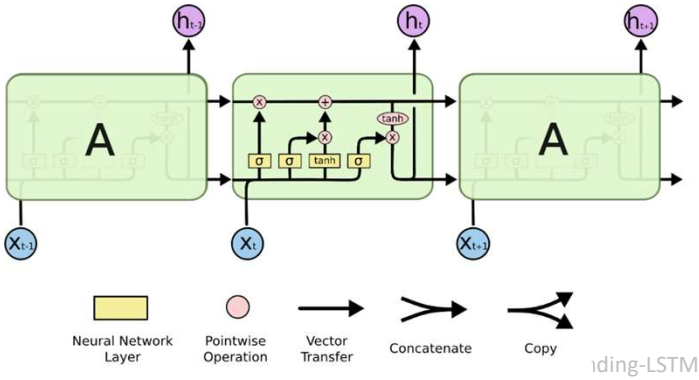
- Difficulty understanding long sequential relation
- Explored in depth by Hochreiter (1991) and Bengio, et al. (1994)

http://colah.github.io/posts/2015-08-Understanding-LSTMs/ 69

69

Long Short Term Memory networks (LSTMs)

- Special kind of RNN
- Capable of learning long-term dependencies
  - Hochreiter & Schmidhuber (1997)

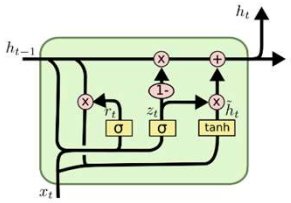


ıding-LSTMs/ 70

70

Gate Recurrent Unit (GRU)

- Cho, et al. (2014)
  - Combines the forget and input gates into a single "update gate."
  - Merges the cell state and hidden state
  - Simpler than standard LSTM



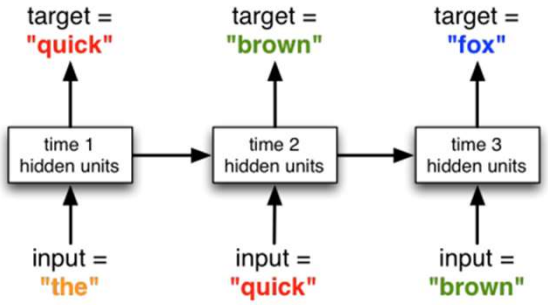
$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$
$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$
$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

http://colah.github.io/posts/2015-08-Understanding-LSTMs/ 71

71

Language Modeling

- Each word is represented as a vector
  - Previous output feeds back in to the network as input



- Dealing with shorter sequence, unseen word issue

72

72

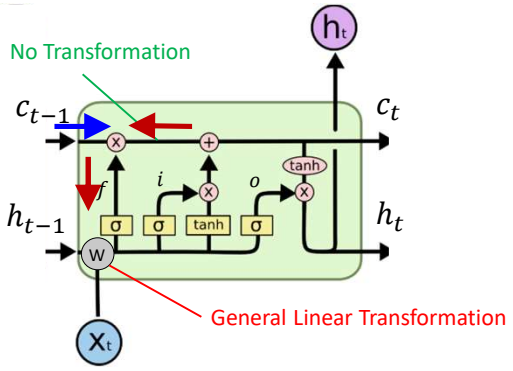
Attention and LSTMs

- We saw something similar in LSTMs: i, f, o nodes learn to weight features.
- They receive a salience gradient during training.

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

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

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

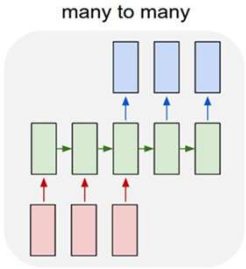


73

Soft Attention

[Bahdanau et al. 2015]

- For Translation
  - "I love coffee" -> "Me gusta el café"
  - "I love coffee" -> "Me gusta el café"
  - "I love coffee" -> "Me gusta el café"

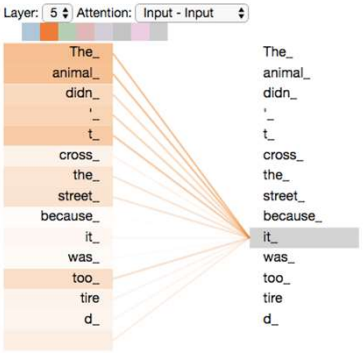


74

74

Self-Attention Example

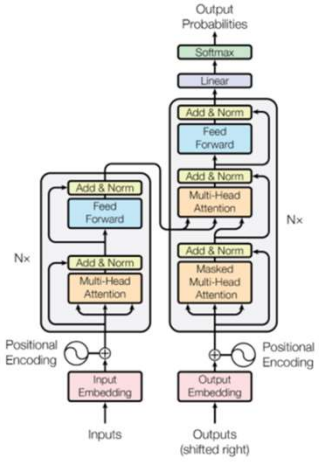
- ”The animal didn't cross the street because it was too tired”
  - What “it” refers too?



"The Animal“ is baked as a part of “it”s representation in encoding 75

75

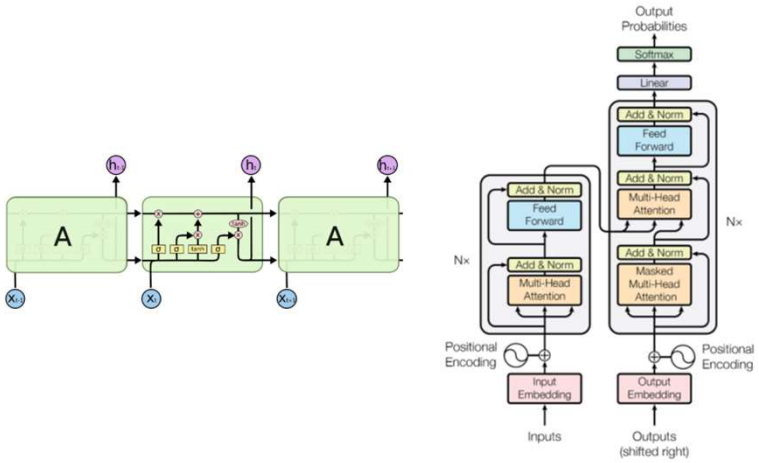
Attention is All you Need!



Jay Alammar https://jalammar.github.io/illustrated-transformer/ 77

77

RNN vs. Attention Based Model



78

78

Outline

Computer Vision (CV)

Transfer Learning (TL) in CV

Hands on TL for CV

Natural Lang. Proc. (NLP)

Transfer Learning (TL) in NLP

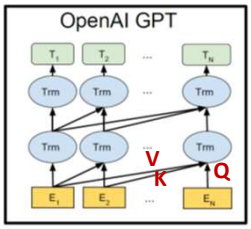
Hands on TL for NLP

79

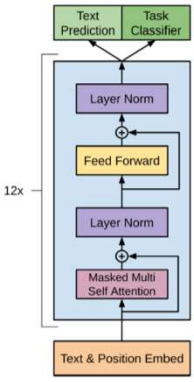
79

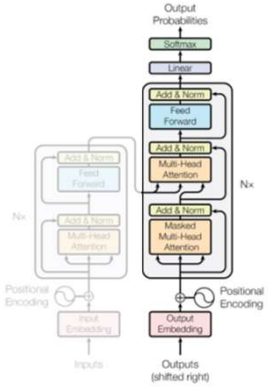
GPT

- Generative Pre-Training (OpenAI) is a transformer-based generator using only a simplified transformer decoder stage:



GPT



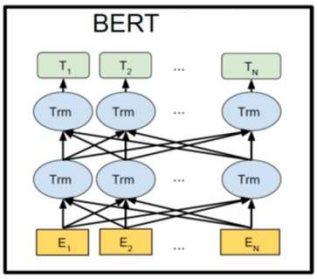


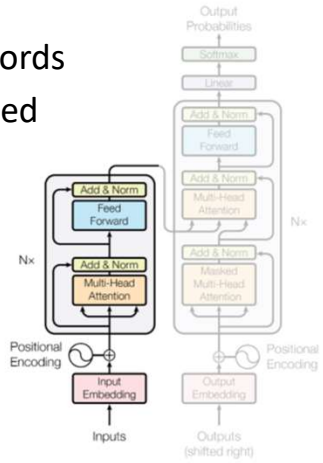
decoder transformer

80

BERT

- BERT is a bidirectional (Transformer encoder) model.
- Word prediction: 15% of input words
- are removed and then re-predicted

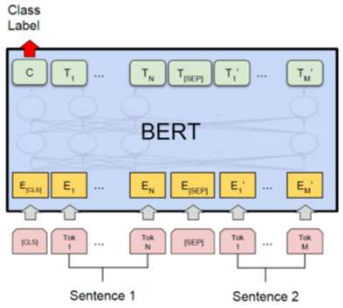




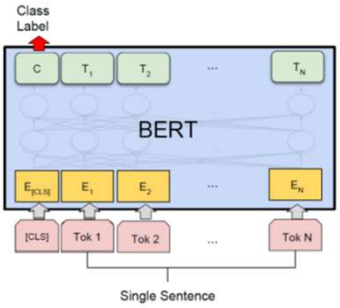
Encoder transformer

81

BERT Task specialization



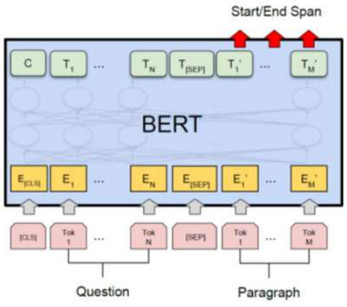
(a) Sentence Pair Classification Tasks: MNLI, QQP, QNLI, STS-B, MRPC, RTE, SWAG



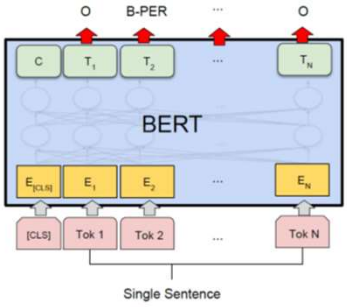
(b) Single Sentence Classification Tasks: SST-2, CoLA

86

BERT Task specialization



(c) Question Answering Tasks: SQuAD v1.1



(d) Single Sentence Tagging Tasks: CoNLL-2003 NER

87

After BERT

- Model that address long sequences
  - Longformer, Big Bird
- Multilingual BERT
- BERT extension to different domain
  - SciBERT, BioBERT, FinBERT, ClinicalBERT
- Making BERT smaller
  - DistillBERT, TinyBERT

88

88

The Landscape

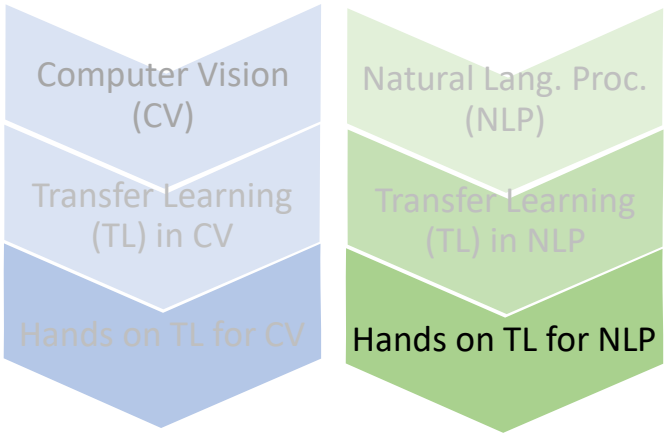
Figure 1: **Parameter counts of several recently released pretrained language models.**

Source: hugging face DistilBERT paper

95

95

Outline



112

112

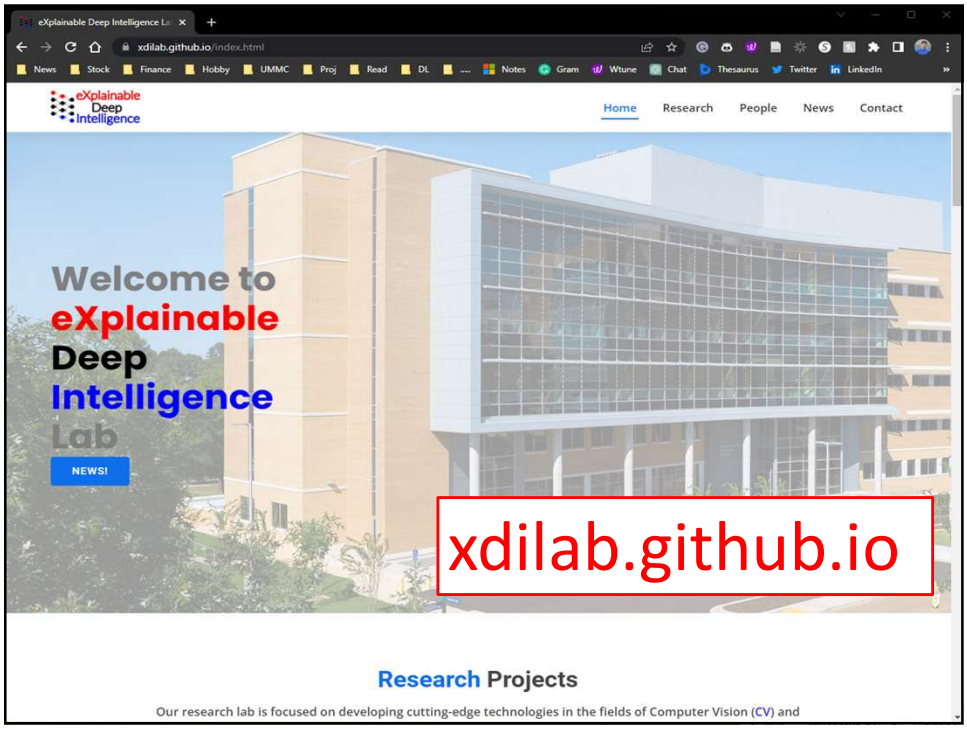
Suicidal Severity Detection

- Gold standard dataset of 500 redditors labeled by psychiatrists following the guidelines outlined in Columbia Suicide Severity Rating Scale (C-SSRS).

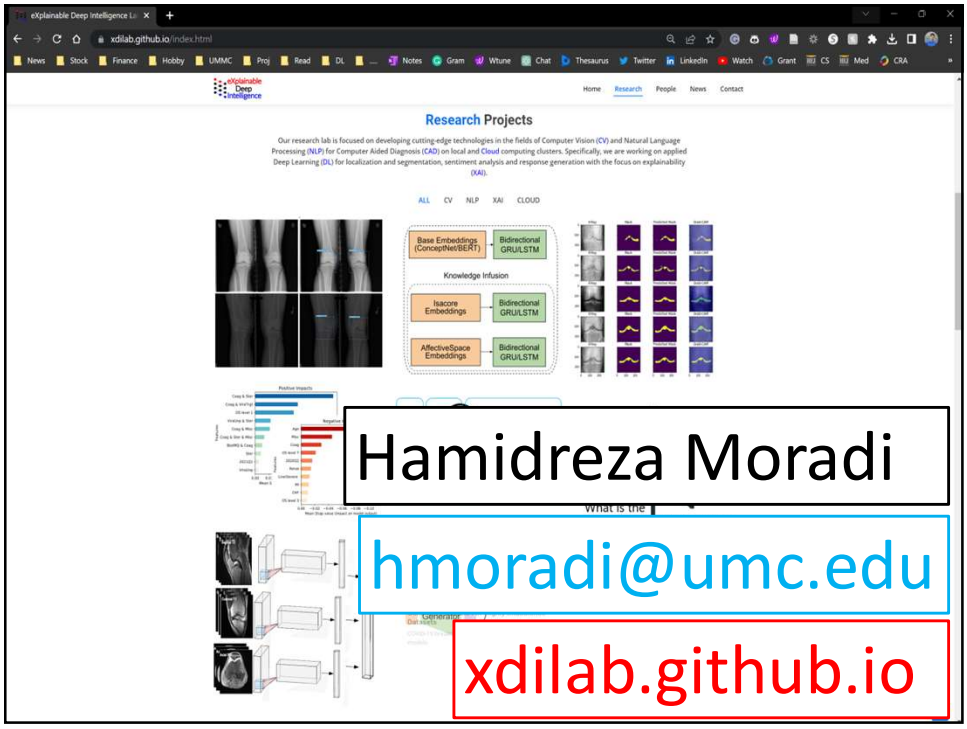
User	Post	Label
user-0	['Its not a viable option, and youll be leaving your wife behind. Youd Pain her beyond comprehensiotat first hand. It can definitely feel hopeless, as you seem to be Tired aware of. Your wife might need to but even 10-15 hours a Asthenia could alleviate a lot of the pressure. In the meantime, get your shit tog excuses, get it done and send it out. Whether you believe in some sort of powerful being or force gc themselves out. This is a big test for you, and youll pull through. Just try to stay as positive as you c	Supportive
user-1	['It can be hard to appreciate the notion that you could meet someone else who will make you happ boyfriend. Your desires are set on him and not much else will make you happy at the moment. But ch loss, this is a proven fact in psychology. Over time, one day you will arrive at a level where you agai comfortable at looking into new relationships.It is certainly uncomfortable dealing with your current stime and patience the Pain will go away and you will get through the difficulties that many US students aand undervalued degrees.These are problems that many of us are facing right now, you are not alone. Ithe same time, but getting through them is what is going to make us stronger, smarter and more emotio before us. ', 'The voice is just a voice. People can praise you, people can hate you but it doesnt chain come up to you and say you are awesome, but does that change you in any way? No. It is a psychowhat	Ideation

https://zenodo.org/record/2667859#.ZB1DS3bMLBk 113

113


xdilab.github.io/index.html
eXplainable Deep Intelligence
Home
Research
People
News
Contact
Welcome to
eXplainable
Deep
Intelligence
Lab
NEWS!
xdilab.github.io
Research Projects
Our research lab is focused on developing cutting-edge technologies in the fields of Computer Vision (CV) and

114


Research Projects
Base Embeddings (ConceptNet/BERT)
Bidirectional GRU/LSTM
Knowledge Infusion
Isacore Embeddings
AffectiveSpace Embeddings
Hamidreza Moradi
hmoradi@umc.edu
xdilab.github.io

115