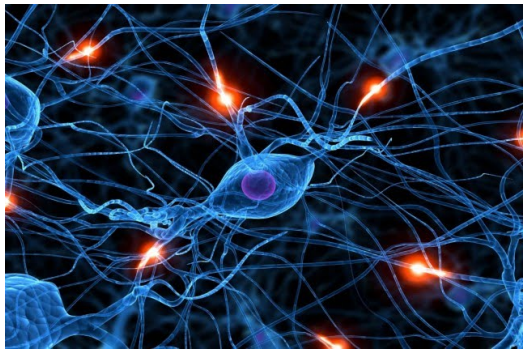
Ph.D. Qualifying Examination

Towards Better Understanding Of Deep Learning With Visualization

Presenter: Haipeng Zeng
Supervised by: Huamin Qu
Department of Computer Science & Engineering
2016.11.10

Introduction-Background

- Deep Learning

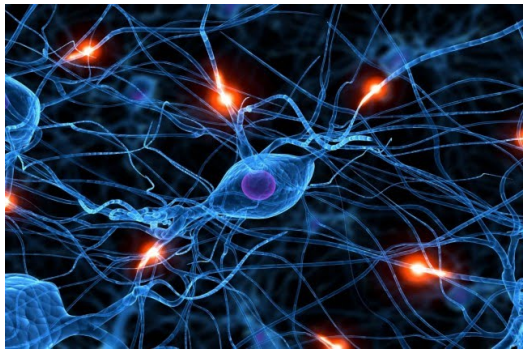


Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction.

(LeCun et al., 2015)

Introduction-Background

- Remarkable Progress



Deep Learning

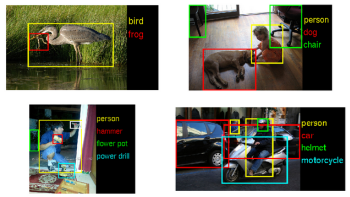
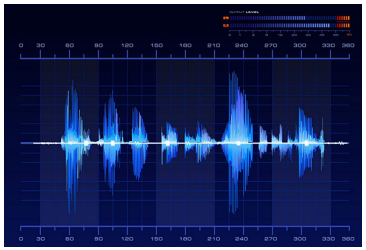
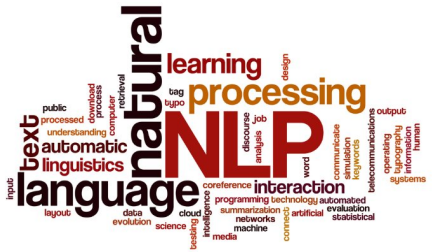


Image Recognition



Speech Recognition



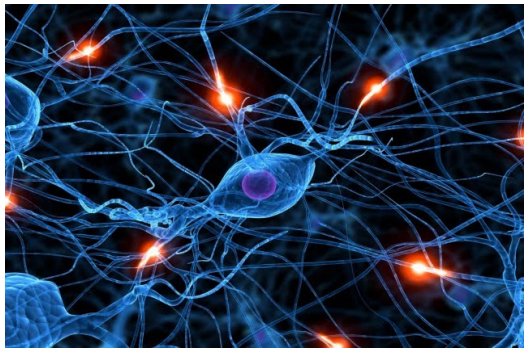
Natural language Processing

. . .

. . .

Introduction-Background

- Black Box
  - No clear understanding of the inner working mechanism
  - Compared with other machine learning models
  - A substantial amount of trial-and-error procedures



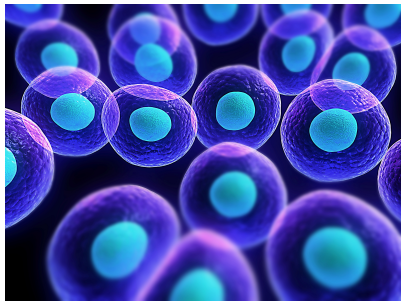
Deep Learning

Visualization

Introduction-Motivation

- A simple analogy

Microscope

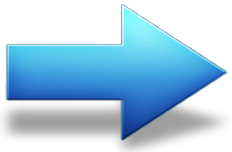


cells

Visualization → **Deep Learning**

Introduction-Motivation

- Visualization on Deep Learning
  - Help understand deep learning models intuitively
  - Help train a better model efficiently
  - Make decisions more interpretable

Visualization  **Deep Learning**

Taxonomy

- A taxonomy based on:
  - the challenges that deep learning faces
  - the purposes that visualization techniques serve

Challenges on Deep Learning:

How a deep learning model works?

How to improve a deep learning model?

Purposes of Visualization:

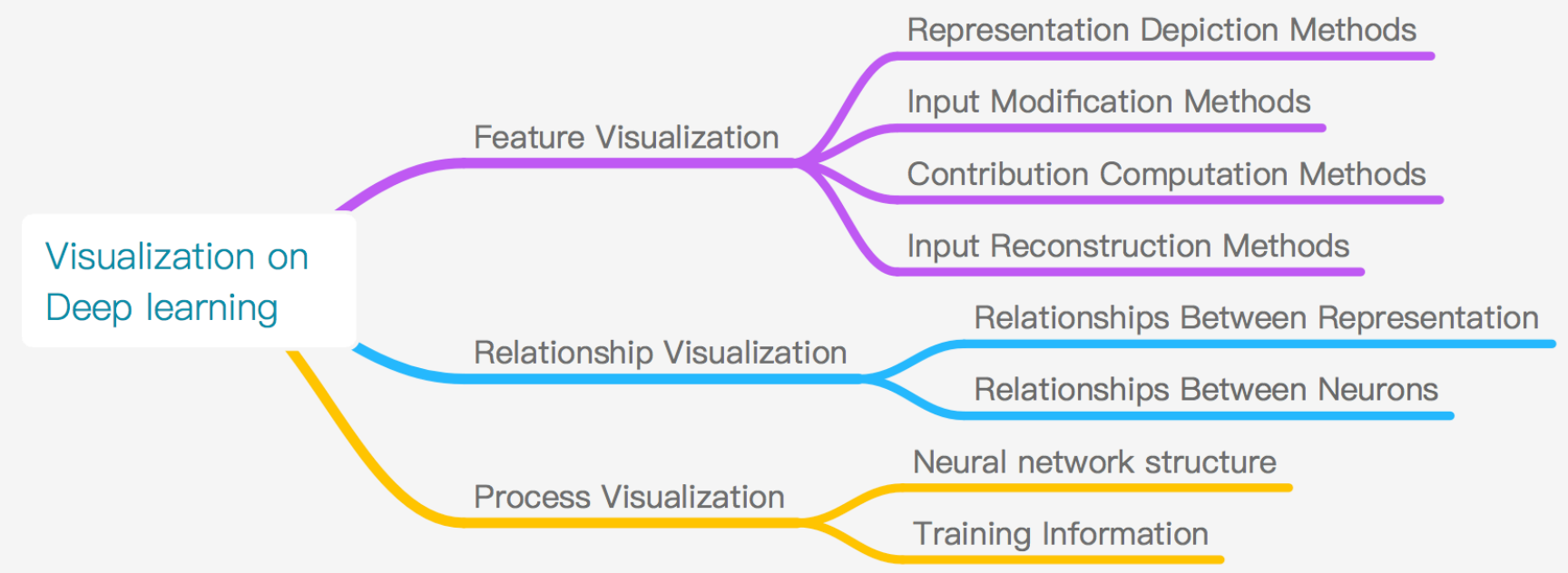
Visualize the features learned by a deep learning model

Visualize the relationships in a deep learning model

Visualize the whole process of a deep learning model

Taxonomy

- A taxonomy based on:
  - the challenges that deep learning faces
  - the purposes that visualization techniques serve



Taxonomy

- The related papers in this survey

	Detail Categories	Related Paper
Feature Visualization	Representation Depiction Methods	(Karpathy, 2014a), (Yosinski et al., 2015), (Karpathy et al., 2015), (Strobelt et al., 2016), etc.
	Input Modification Methods	(Zeiler & Fergus, 2014), (Zhou et al., 2014), (Girshick et al., 2014), etc.
	Contribution Computation Methods	(Zeiler & Fergus, 2014), (Simonyan et al., 2013), (Springenberg et al., 2014), (Bach et al., 2015), (Zhou et al., 2015), (Bahdanau et al., 2014), (Socher et al., 2013), (Hermann et al., 2015), (Goyal et al., 2016), etc.
	Input Reconstruction Methods	(Long et al., 2014), (Erhan et al., 2009), (Simonyan et al., 2013), (Alexander et al., 2015), (Mahendran & Vedaldi, 2015), (Yosinski et al., 2015), (Mahendran & Vedaldi, 2016), (Dosovitskiy & Brox, 2015), etc
Relationship Visualization	Relationships Between Representations	(Maaten & Hinton, 2008), (Cho et al., 2014), (Karpathy, 2014b), (Rauber et al., 2016), etc.
	Relationships Between Neurons	(Liu et al., 2016), (Rauber et al., 2016), etc.
Process Visualization	Neural Network Structure (Model)	(Karpathy, 2014a), (Yosinski et al., 2015), (Smilkov et al., 2015), (Harley, 2015), (Chung et al., 2016), (Liu et al., 2016), (Bruckner, 2014), (Google, 2015), etc.
	Training Information (Data)	(Google, 2015), (Bruckner, 2014), (Smilkov et al., 2015), (Skymind, 2013), (Chung et al., 2016), etc.

Outline

- Deep Learning Models
- Feature Visualization
- Relationship Visualization
- Process Visualization
- Conclusion and Future Work

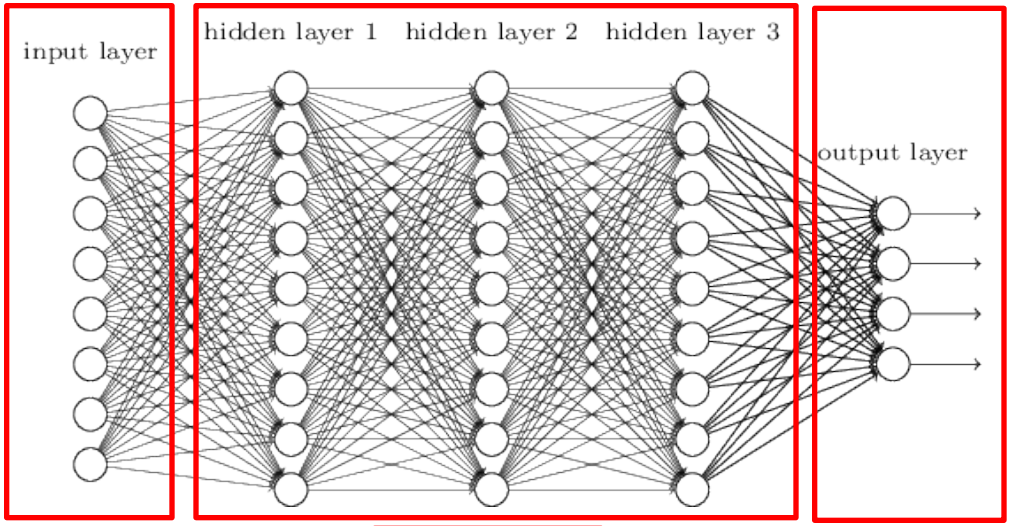
Outline

- Deep Learning Models
  - Deep Neural Network (DNN)
  - Convolutional Neural Network (CNN)
  - Recurrent Neural Network (RNN)
- Feature Visualization
- Relationship Visualization
- Process Visualization
- Conclusion and Future Work

Deep Learning Models - DNN

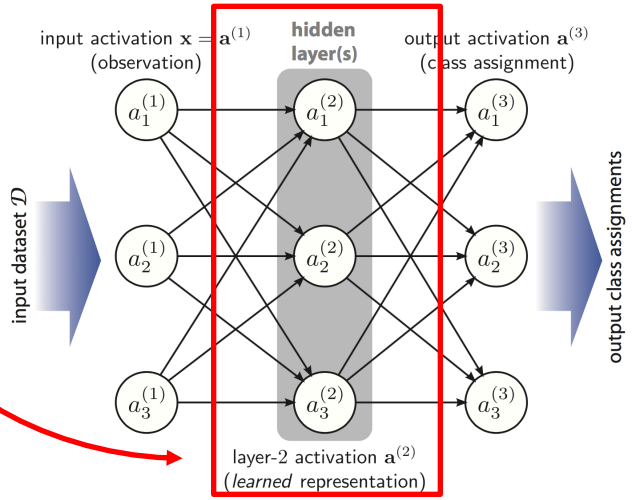
THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

- Input layer
- Output layer
- Hidden layers
- Neurons (units)



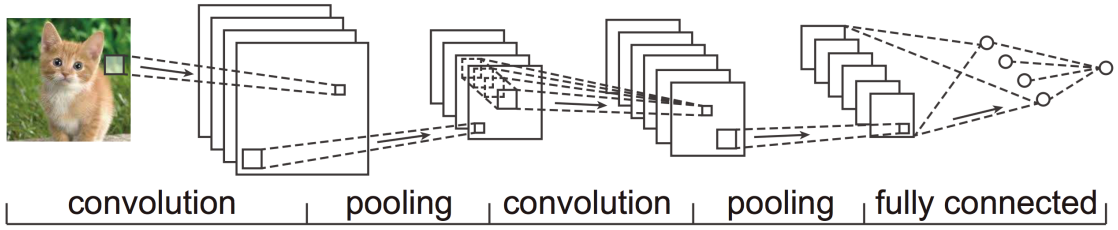
- Activation
- Representation

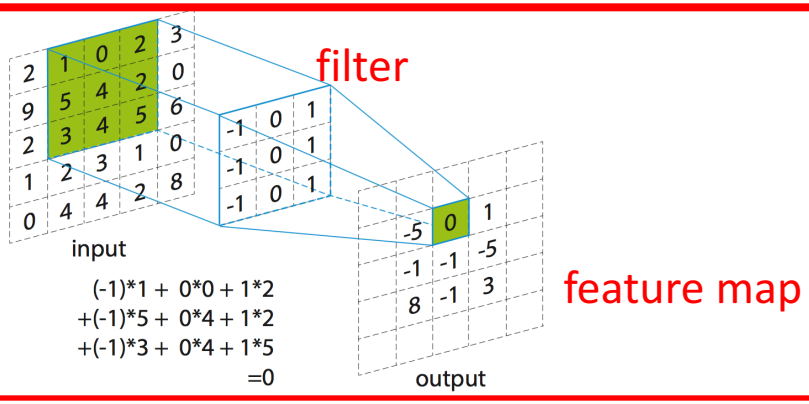
- Backpropagation

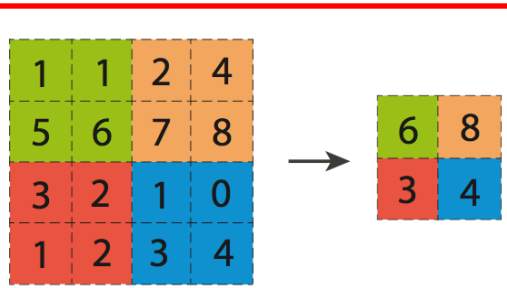


Deep Learning Models - CNN

- Architecture
  - Convolution layer
  - Pooling layer
  - Fully connected layer
- Convolution
  - Filter (kernel)
  - Feature map (Activation Map)
- Max-pooling
  - Translation invariance

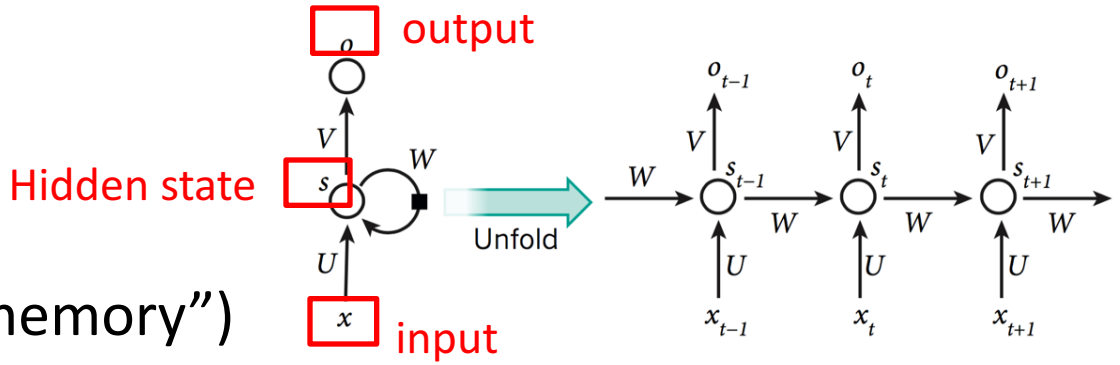




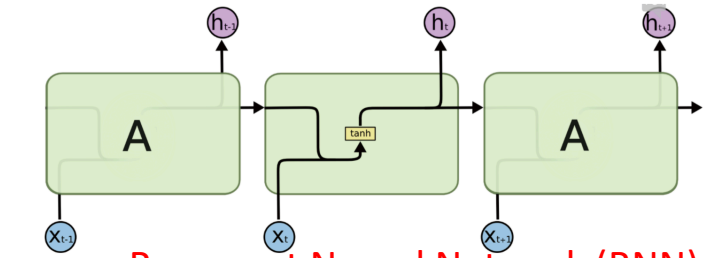


Deep Learning Models - RNN

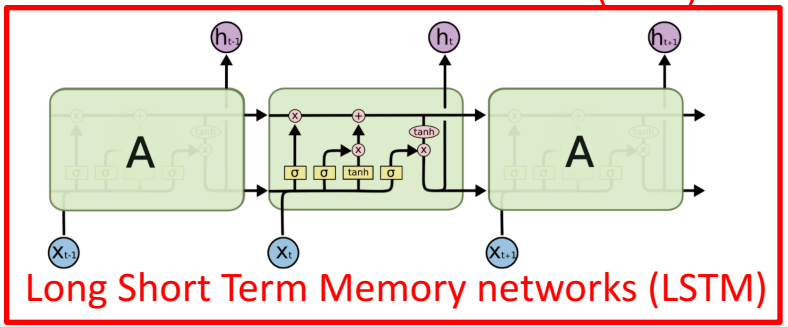
- Input
- Output
- Hidden state (“memory”)



- LSTMs
  - Input gate
  - Forget gate
  - Output gate



Recurrent Neural Network (RNN)



Long Short Term Memory networks (LSTM)

Outline

- Deep Learning Models
- Feature Visualization
  - Representation Depiction Methods

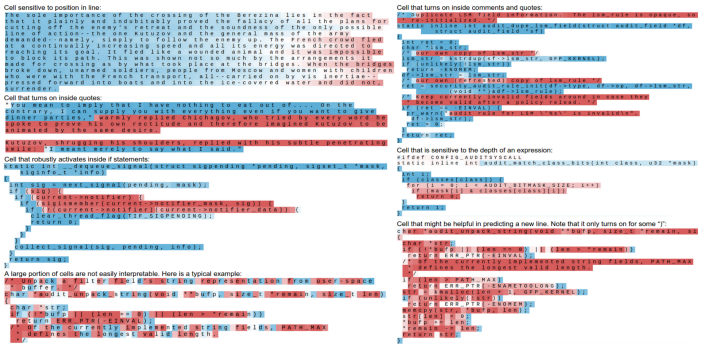
Representation Depiction Methods refer to those methods that visually depict representation directly.

PS.
A representation is a vector of activations in a hidden layer.

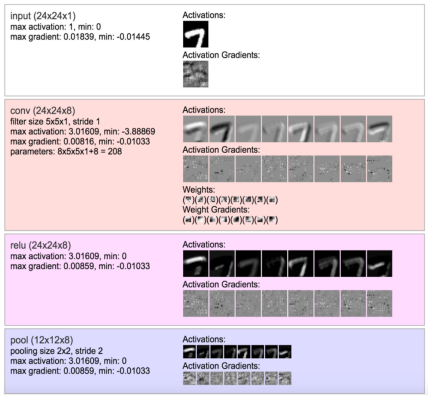
Representation Depiction Methods

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

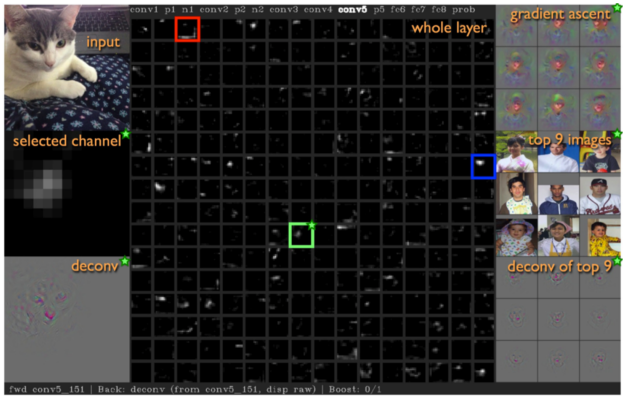
- CNNs
  - Activation map (Feature map)
- LSTMs
  - Hidden state



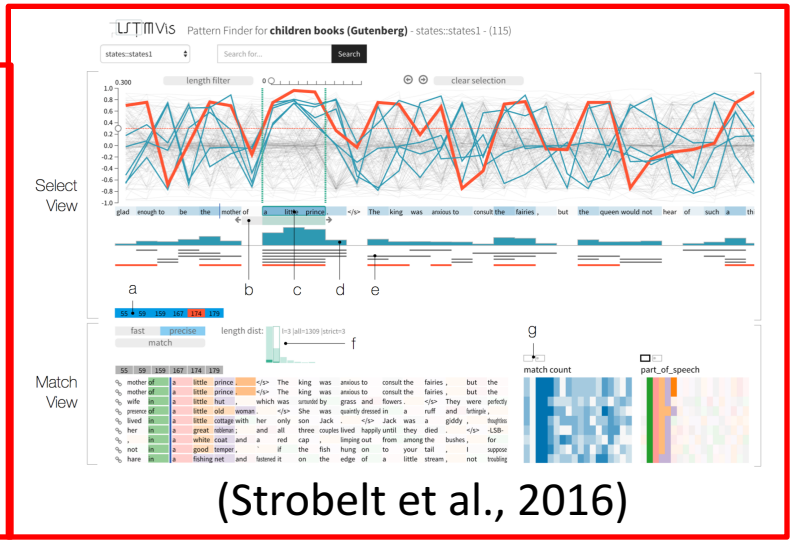
(Karpathy et al., 2015)



(Karpathy, 2014a)



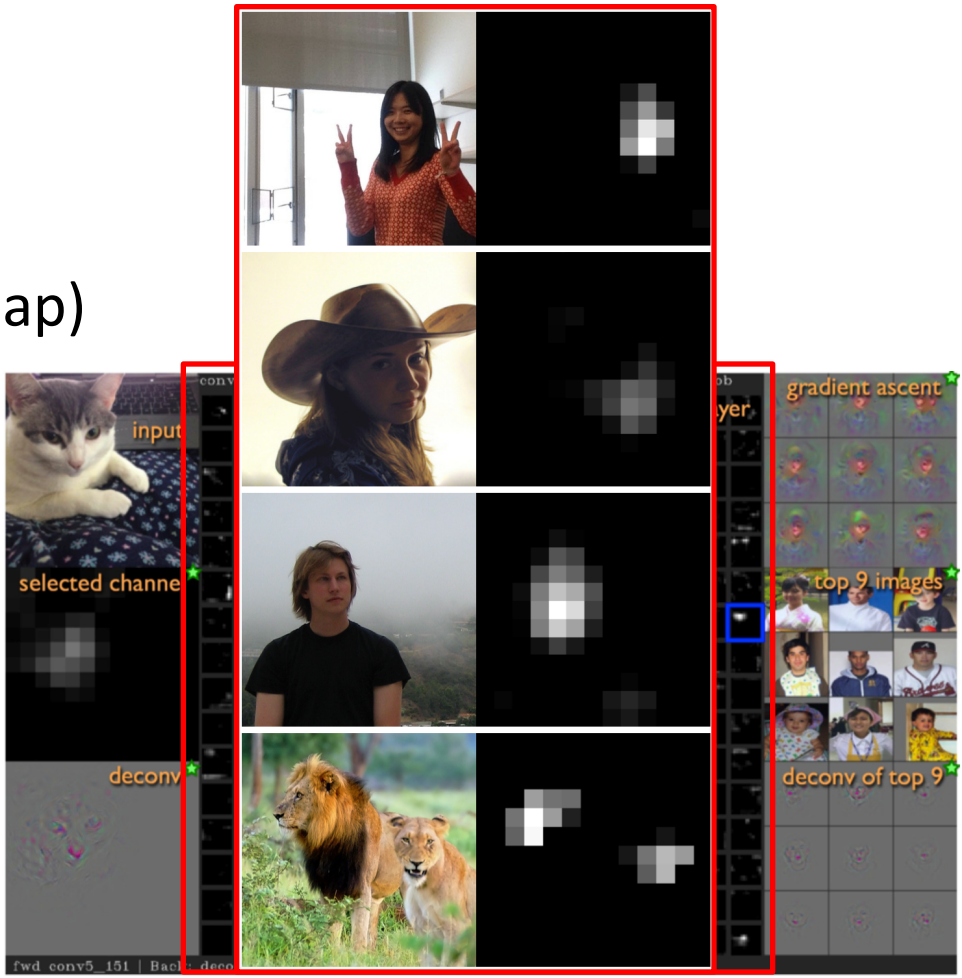
(Yosinski et al., 2015)



(Strobelt et al., 2016)

Representation Depiction Methods

- CNNs
  - Deep visualization system
  - Activation map (Feature map)



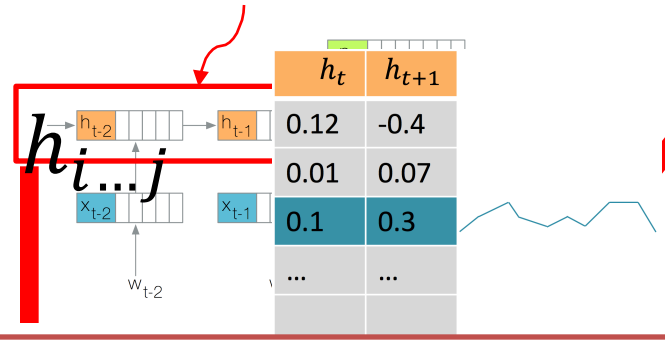
(Yosinski et al., 2015)

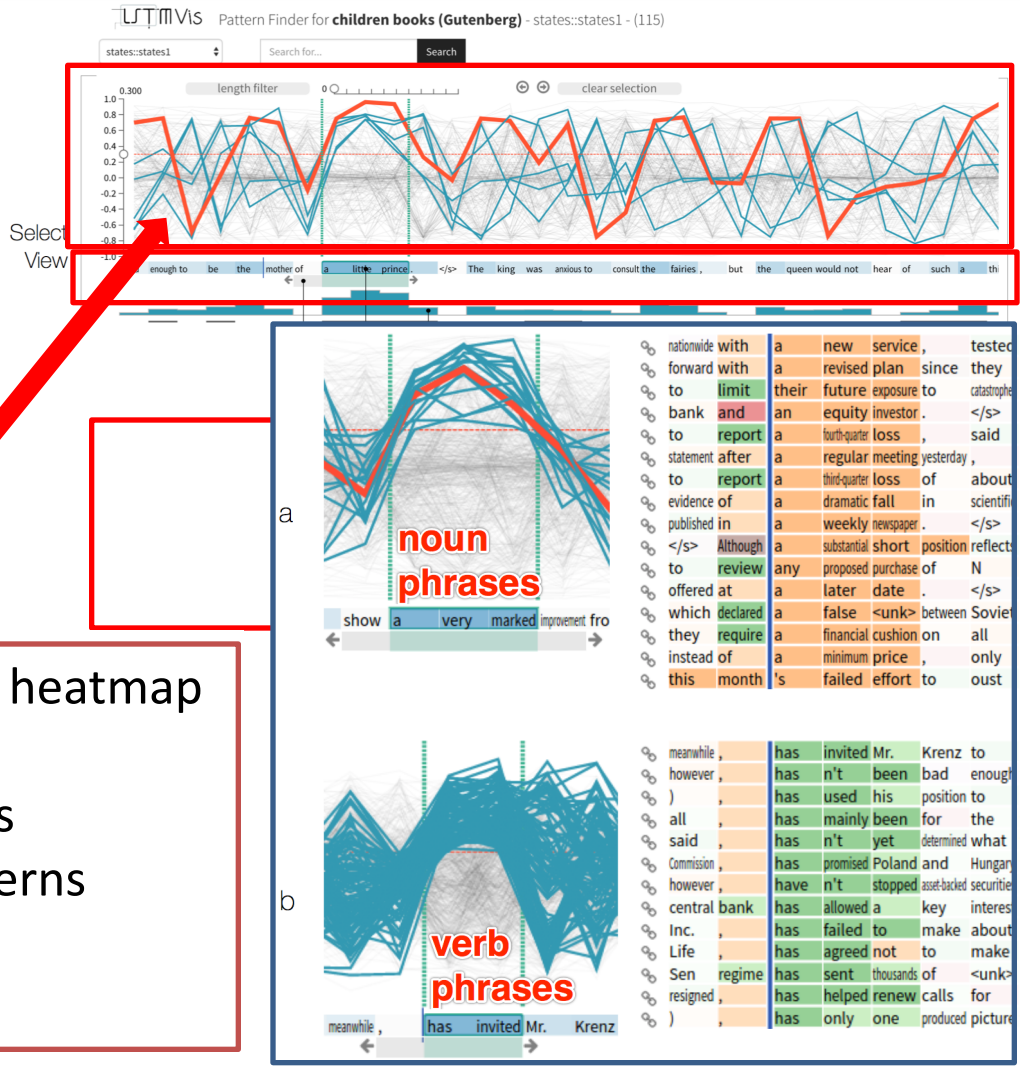
Visualization: bitmap
Pros:
+ simple implement and informative
+ support webcam input
Cons:
- scalability problem
- Hard to explain in some cases

Representation Depiction Methods

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

- LSTMs
  - Hidden state





Visualization: parallel coordinates + heatmap

Pros:

+ show the hidden state dynamics

+ match similar hidden state patterns

Cons:

- visual clutter

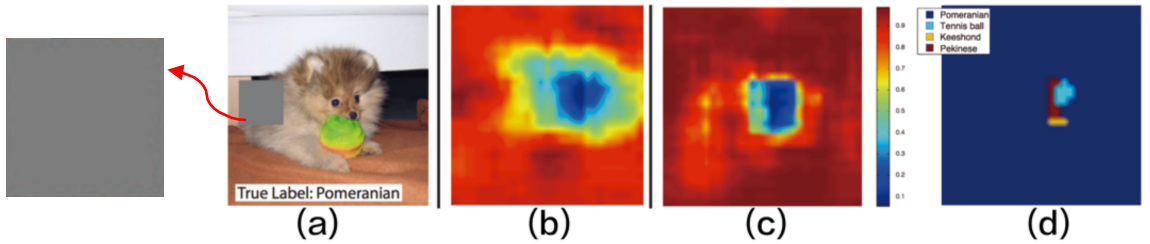
Outline

- Deep Learning Models
- Feature Visualization
  - Representation Depiction Methods
  - Input Modification Methods

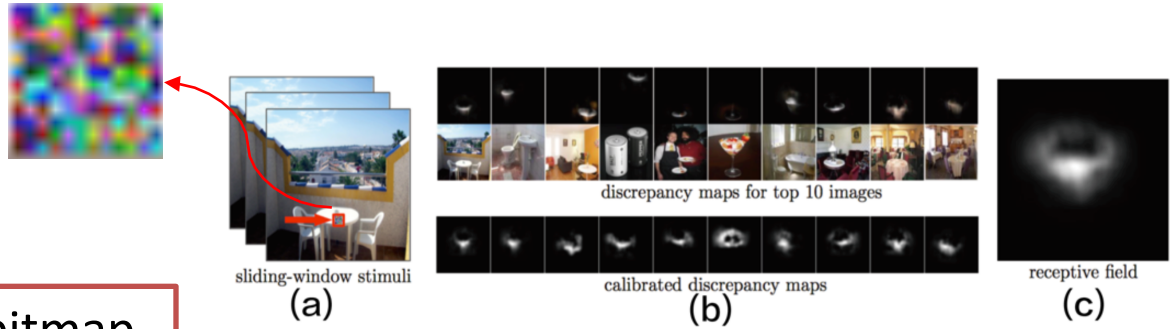
Input Modification Methods refer to those methods where we modify the input and then measure the changes of the output or activations in hidden layers.

Input Modification Methods

- A gray square
  - mono colored
- A randomized patch



(Zeiler & Fergus, 2014)



(Zhou et al., 2014)

Visualization: heatmap+bitmap
Pros:
+ know which parts are important
Cons:
- affected by the square or patch

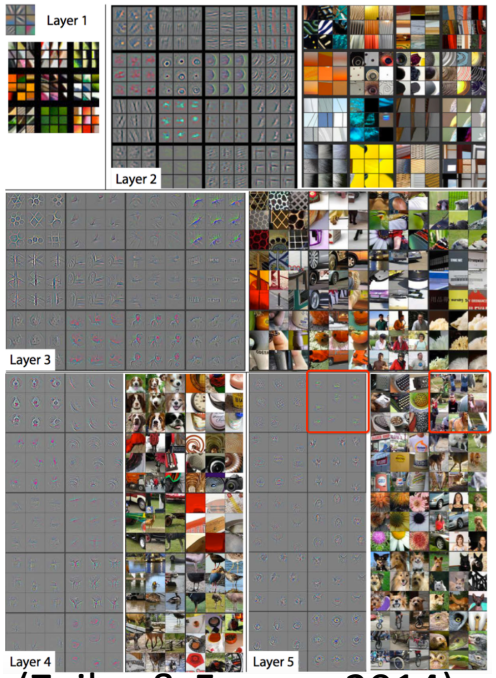
Outline

- Deep Learning Models
- Feature Visualization
  - Representation Depiction Methods
  - Input Modification Methods
  - Contribution Computation Methods

Contribution Computation Methods refer to those methods that compute the contributions of input to the result.

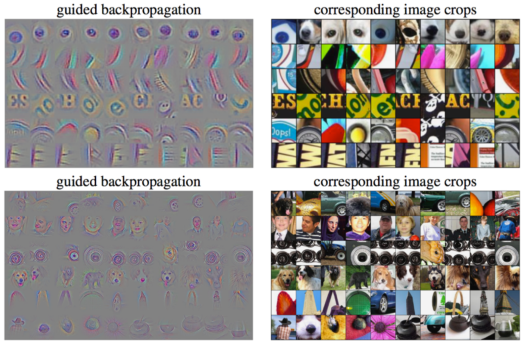
Contribution Computation Methods

- Deconvolutional network
- Backpropagation
- Guide backpropagation
- Relevance propagation
- Class activation map

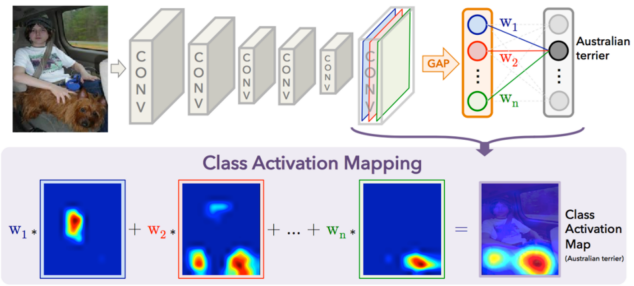


(Zeiler & Fergus, 2014)

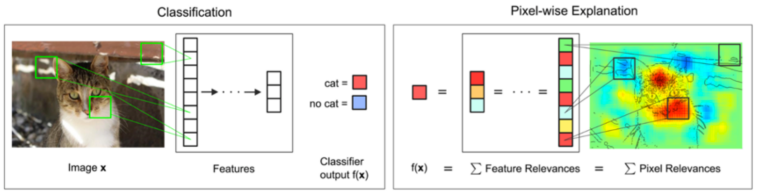
(Simonyan et al., 2013)



(Springenberg et al., 2014)



(Zhou et al., 2015)



(Bach et al., 2015)

Contribution Computation Methods

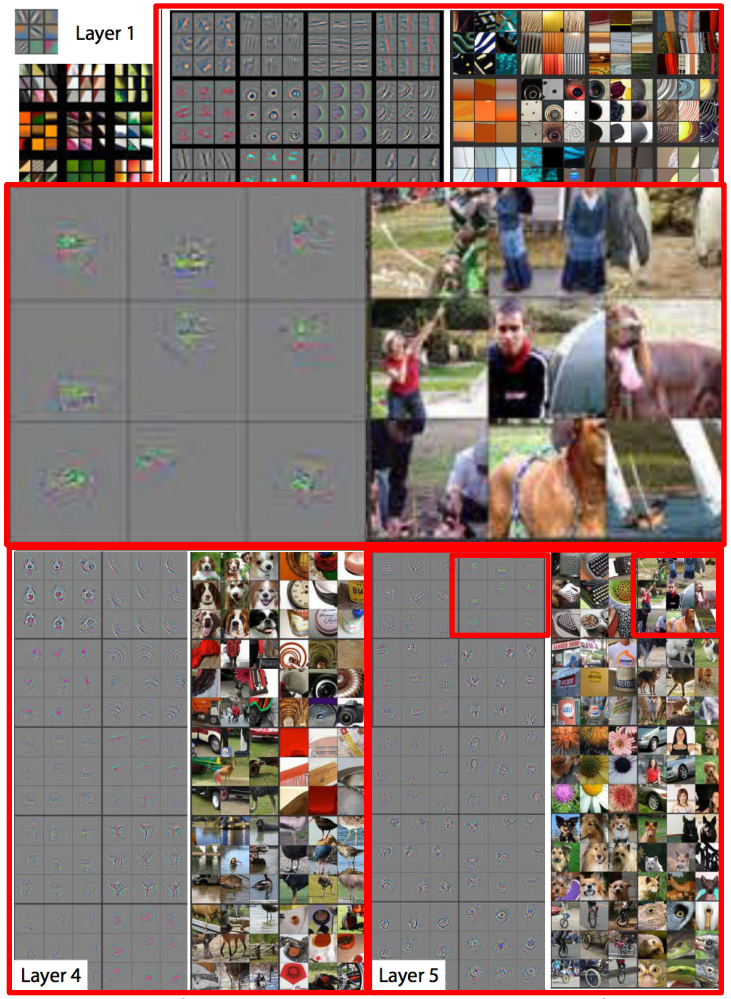
- Deconvolutional network

Layer 2

Remarkable results:
Low layers detect low features: edge, color, etc.
High layers detect high features: object, etc.
has more complex invariances.

Visualization: bitmap
Pros:
+ know each pixel's contribution
+ provide a non-parametric view of invariance
Cons:
- cannot visualize the joint activity in a layer

pose variation, e.g. keyboards, dogs

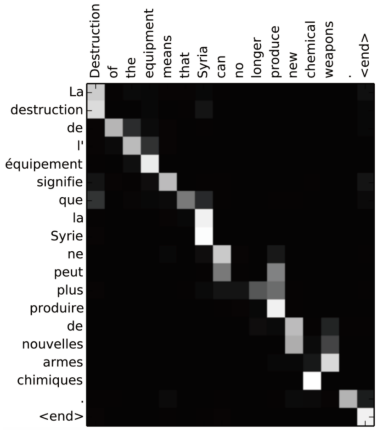


(Zeiler & Fergus, 2014)

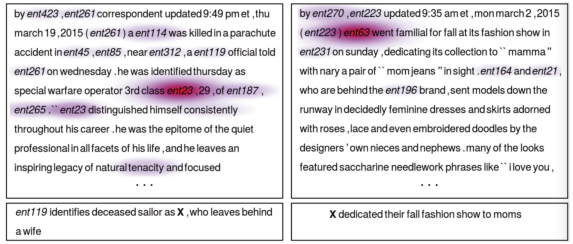
Contribution Computation Methods

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

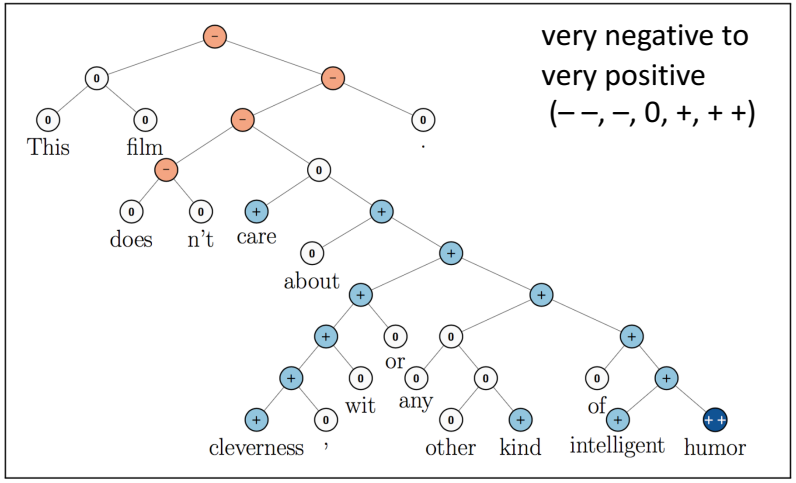
Visualization on text:
- Tree structure
- Matrix
- Heatmap



(Bahdanau et al., 2014)



(Hermann et al., 2015)



(Socher et al., 2013)

(Goyal et al., 2016)

Outline

- Deep Learning Models
- Feature Visualization
  - Representation Depiction Methods
  - Input Modification Methods
  - Contribution Computation Methods
  - Input Reconstruction Methods

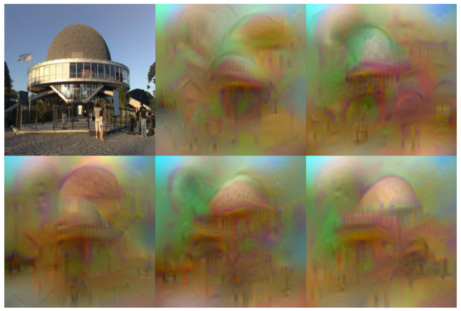
Contribution Computation Methods refer to the methods that reconstruct the input based on representations in the network.

Input Reconstruction Methods

- Gradient
- Replacement
- A generative network

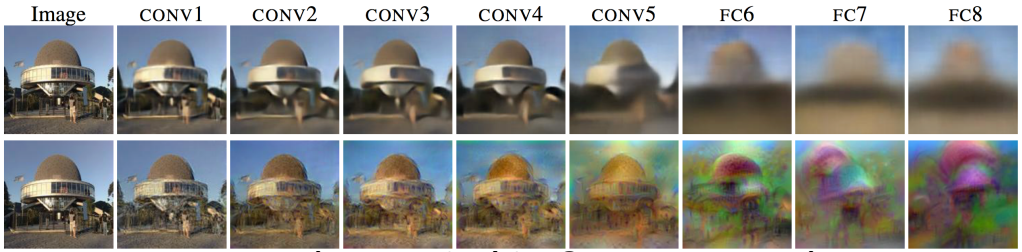
(Simonyan et al., 2013)

(Alexander et al., 2015)



(Mahendran & Vedaldi, 2015)

(Long et al., 2014)



(Dosovitskiy & Brox, 2015)

Input Reconstruction Methods

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

- Gradient
  - Activation maximization
  - Code inversion

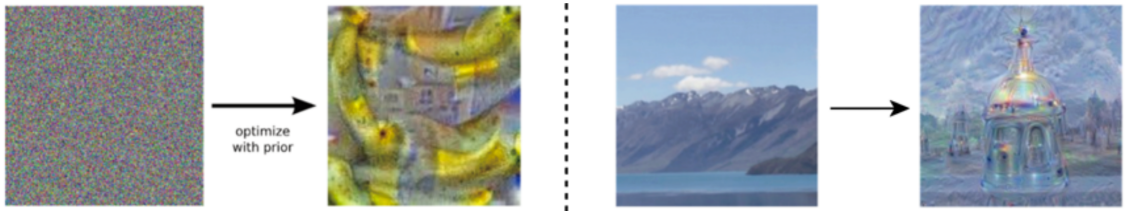
Activation maximization aims to find an image that maximally activates the neuron of interest.

Visualization: image
Pros:
+ generate a artificial input image
Cons:
- need natural-image priors
- hard to get global optimization
- some repeats

(Simonyan et al., 2013)

Input Reconstruction Methods

- Gradient
  - Activation maximization
  - Code inversion



Deep Dream (inceptionism)

(Alexander et al., 2015)

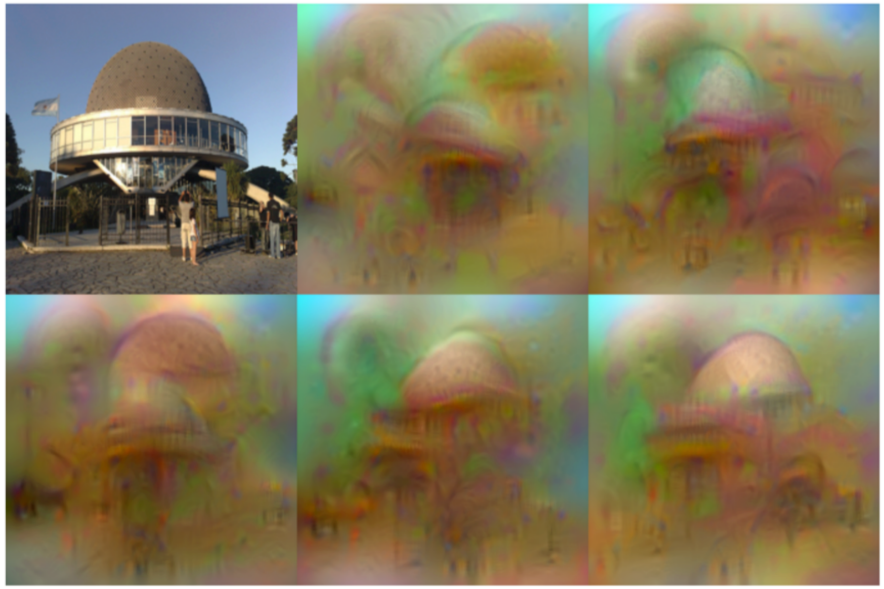
Input Reconstruction Methods

- Gradient
  - Activation maximization
  - Code inversion

Code inversion aims to synthesize an image starting from the encoded image representation.

Visualization: image
Pros:
1. generate artificial input images

Cons:
1. need natural-image priors
2. inversion is not unique



(Mahendran & Vedaldi, 2015)

Feature Visualization

Category	Pros	Cons
Representation Depiction Methods	the most direct way to visualize representations, etc.	scalability problem; visual clutter; may be difficult to be explained in some cases, etc.
Input Modification Methods	easy to know which parts are important, etc.	just crops of input images; don't know detail pixels' contributions, etc.
Contribution Computation Methods	Intuitively get how inputs contribute to results, etc.	need natural-image priors, not clear how to evaluate the quality of a heatmap, etc.
Input Reconstruction Methods	generate artificial input images , etc.	need natural-image priors, images not unique, hard to get global optimization, etc.

Outline

- Deep Learning Models
- Feature Visualization
- Relationship Visualization
  - Relationships between representations
  - Relationships between neurons
- Process Visualization
- Conclusion and Future Work

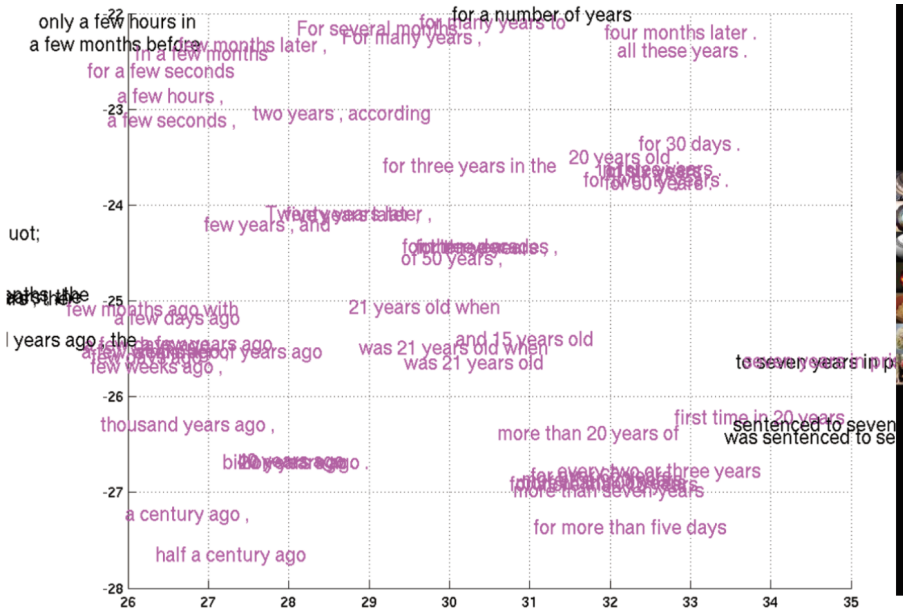
Relationships Between Representations

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

- Scatter-plot visualization
  - T-SNE projection

It projects high dimensional vectors into 2D plane and preserve neighborhoods and clusters.

Text



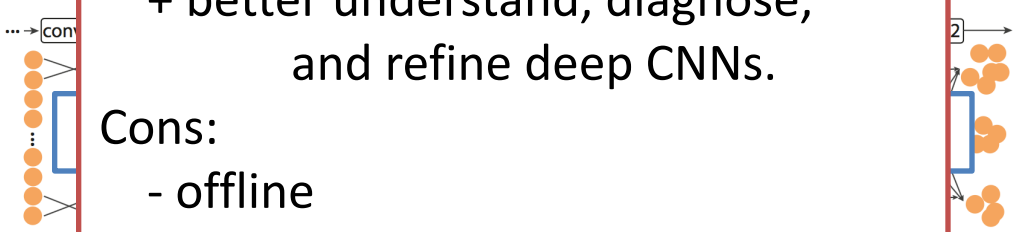
(Cho et al., 2014)

Image

(Karpathy, 2014b)

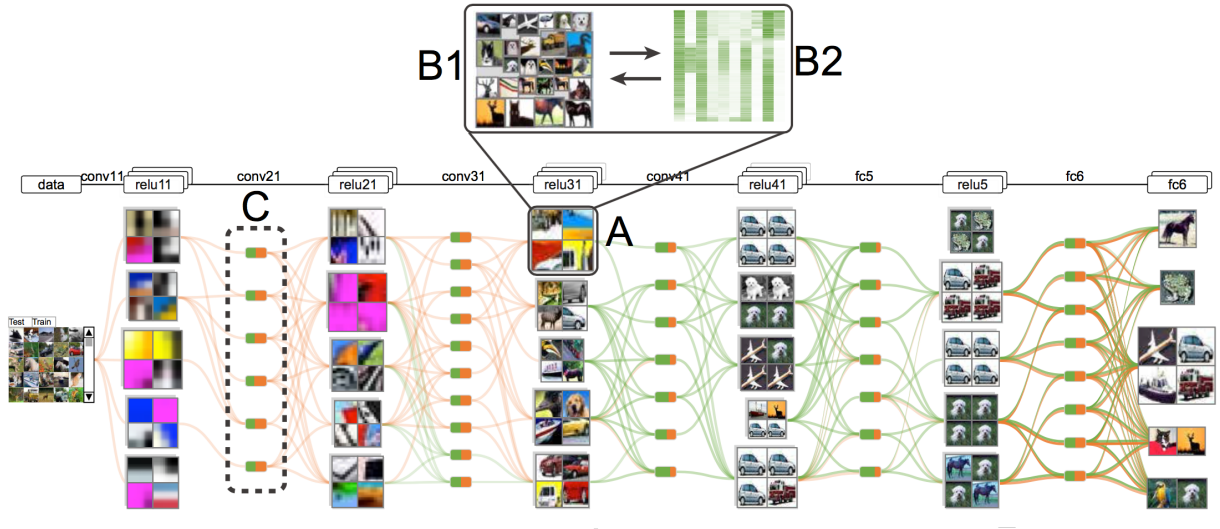
Relationships Between Neurons

- Directed acyclic graph
- A hybrid visualization
  - rectangle packing
  - matrix visualization
  - a biclustering-based edge bundling



Pros:
+ better understand, diagnose, and refine deep CNNs.

Cons:
- offline
- DAG, cannot support RNNs, etc.



(Liu et al., 2016)

Relationship Visualization

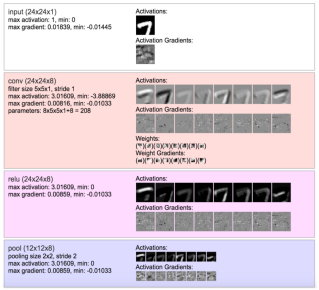
Category	Pros	Cons
Scatter-plot visualization (t-SNE, MDS, etc.)	intuitive; visualize relationships between representation; visualize relationships between neurons; easy to extend to other models, etc.	visual clutter; Projection result change, etc.
DAG-based visualization (clustering, etc.)	intuitive; capture whole picture; visualize relationships between neurons; show relationships between layers, etc.	visual clutter; offline; can not apply to RNNs, etc.

Outline

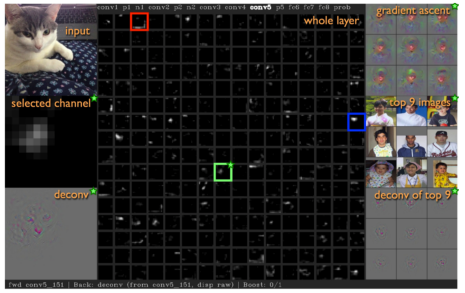
- Deep Learning Models
- Feature Visualization
- Relationship Visualization
- Process Visualization
  - Neural Network Structure
  - Training Information
- Conclusion and Future Work

Neural Network Structure

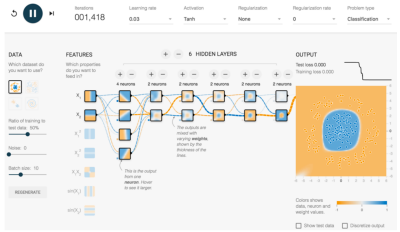
- Grid-Based Diagrams
- Node-Link Diagrams
- Block-Link Diagrams



(Karpathy, 2014a)

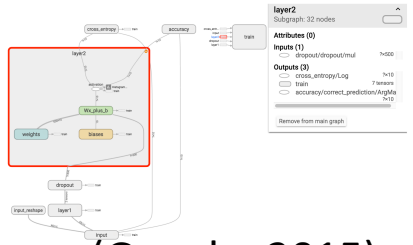


(Yosinski et al., 2015)

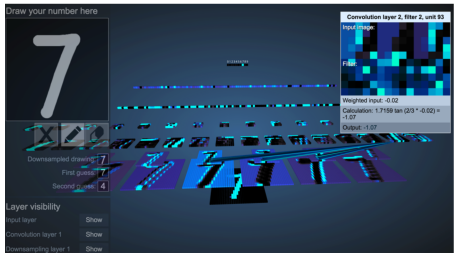


(Smilkov et al., 2015)

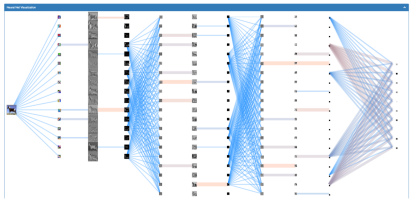
(Bruckner, 2014)



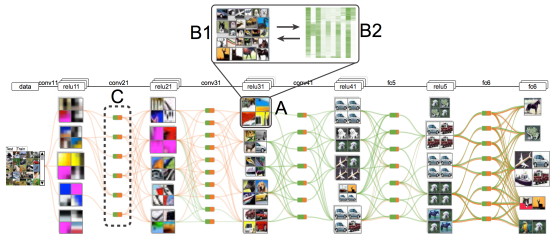
(Google, 2015)



(Harley, 2015)



(Chung et al., 2016)



(Liu et al., 2016)

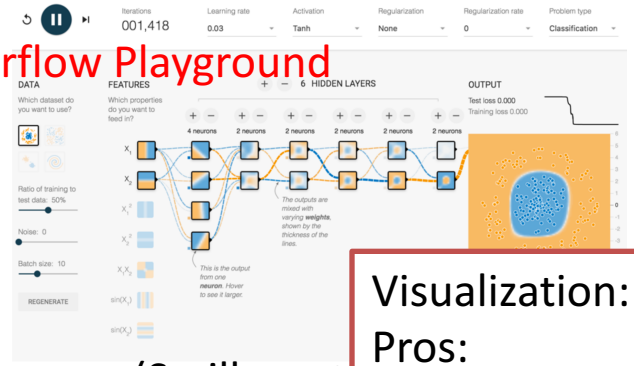
Neural Network Structure

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

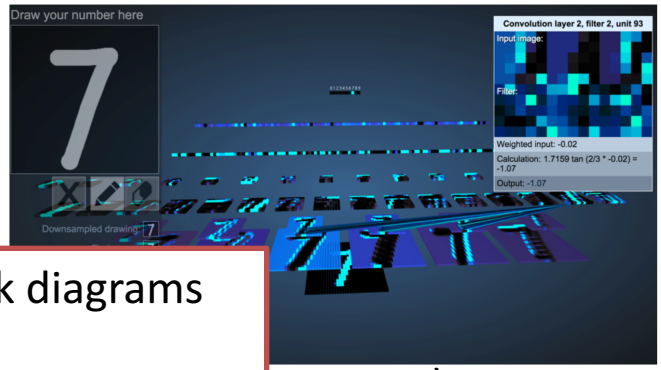
- Node-Link Diagrams

CNNVis:
Aggregate related layers and cluster neurons; bicluster edge bundling

Tensorflow Playground



(Smilkov et

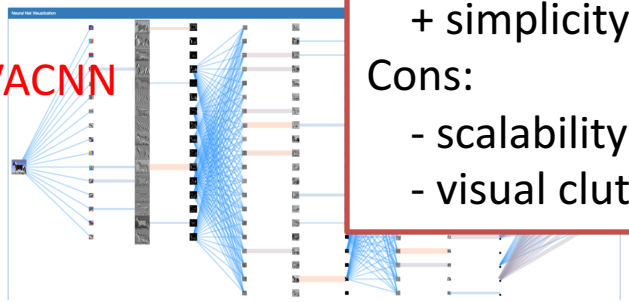


CNN-3D

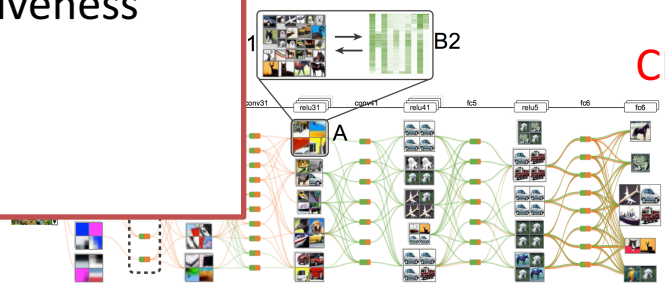
ey, 2015)

Visualization: Node-link diagrams
Pros:
+ give an overview
+ simplicity and intuitiveness
Cons:
- scalability
- visual clutter

ReVACNN



(Chung et al., 2016)

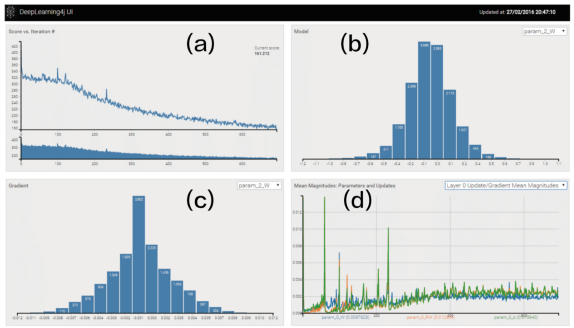


CNNVis

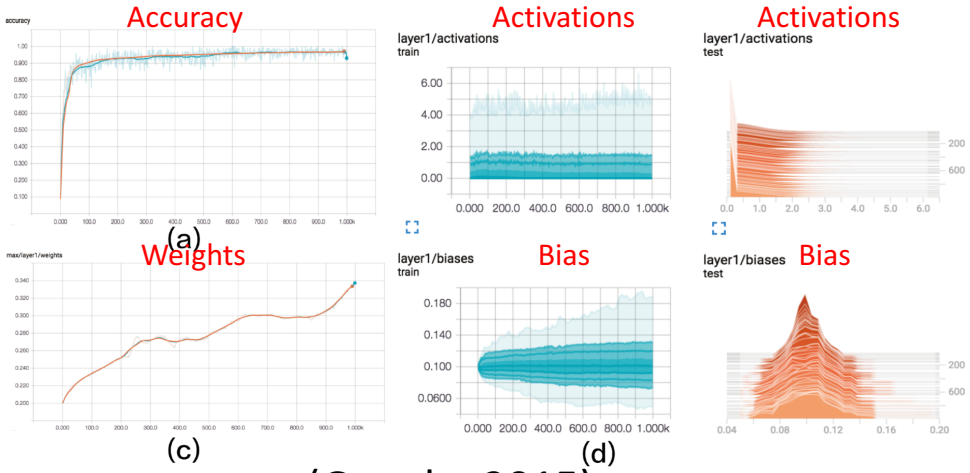
(Liu et al., 2016)

Training Information

- Visualization after Training
- Visualization during Training

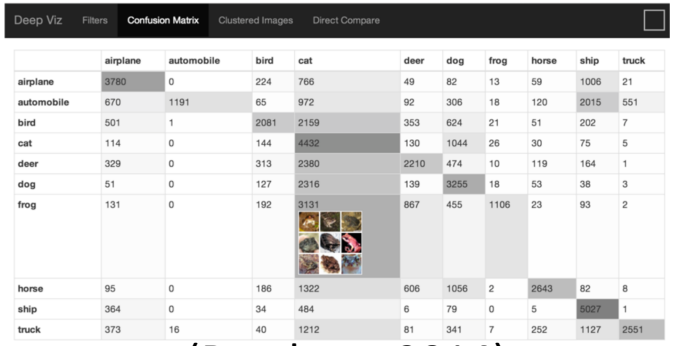


(Skymind, 2013)



(Google, 2015)

Visualization:
line chart, bar chart, histogram, matrix, etc.
Pros:
+ intuitive
Cons:
- trival design



	airplane	automobile	bird	cat	deer	dog	frog	horse	ship	truck
airplane	3780	0	224	766	49	82	13	59	1006	21
automobile	670	1191	65	972	92	306	18	120	2015	551
bird	501	1	2081	2159	353	624	21	51	202	7
cat	114	0	144	4432	130	1044	26	30	75	5
deer	329	0	313	2380	2210	474	10	119	164	1
dog	51	0	127	2316	139	3255	18	53	38	3
frog	131	0	192	3131	867	455	1106	23	93	2
horse	95	0	186	1322	606	1056	2	2643	82	8
ship	364	0	34	484	6	79	0	5	5027	1
truck	373	16	40	1212	81	341	7	252	1127	2551

(Bruckner, 2014)

Process Visualization

Data Category	Data types	Visualization
input data	images, text, etc.	bitmap, heatmap, t-SNE, tree structure, matrix, etc.
hidden layer data	activation maps, filters, hidden state, etc.	bitmap, heatmap, parallel coordinates, t-SNE, etc.
output data	loss function, accuracy, classification results, etc.	line chart, confusion matrix, etc.
parameters	weights, biases, etc.	line chart, histogram, bar chart, etc.
hyper-parameters	the number of layers, the number of neurons in each layer, learning rate, batch size, etc.	line chart, bar chart, etc.

Outline

- Deep Learning Models
- Feature Visualization
- Relationship Visualization
- Process Visualization
- Conclusion and Future Work

Conclusion

- Integrate visualization with deep learning
  - Feature Visualization
  - Relationship Visualization
  - Process Visualization

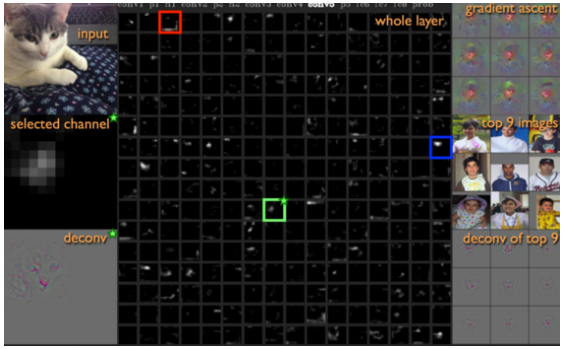
- Advantages
  - Understand what features learned by deep learning models
  - Grasp the inner working mechanism of deep learning models
  - Facilitate people to design and train better deep learning models
  - Make deep learning models more understandable and accessible to people

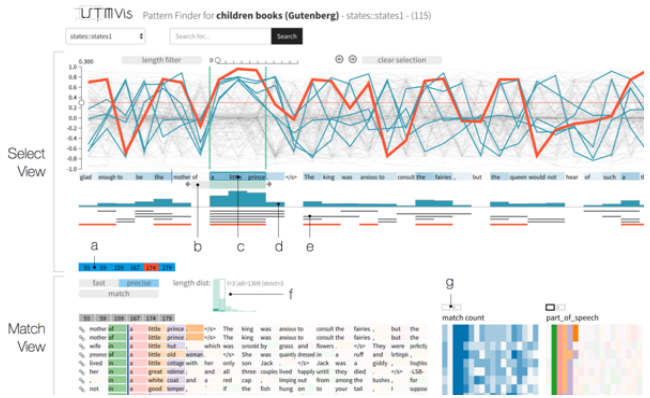
Future Work

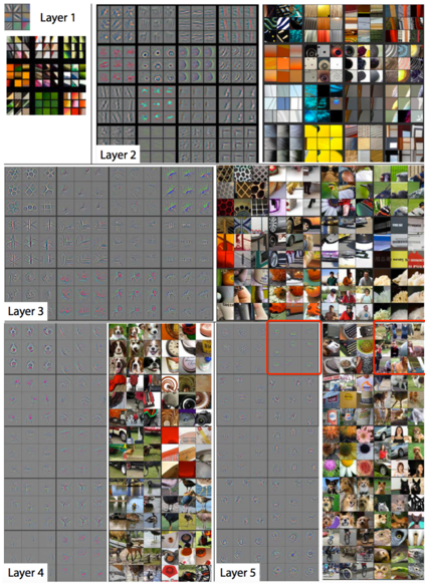
- New designs are needed.
  - Solve scalability issues
  - Provide informative insight
- A visual analysis system is needed.
  - Instant and iterative feedback
  - A friendly user interface
  - Smooth interaction
- Extend visualization techniques to other data and models.
- Combine visualization with deep learning models on different applications.

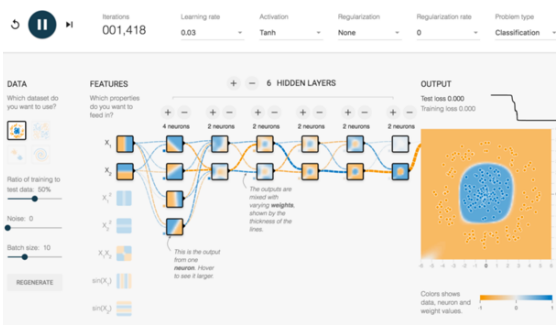
Ph.D. Qualifying Examination

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系









Thank you!