Today: Outline

- Model Evaluation
- A Dropout Variant
- Model Explainability

- Reminders: *Pre-lec Material 2,*
  due: Friday, Jun 4

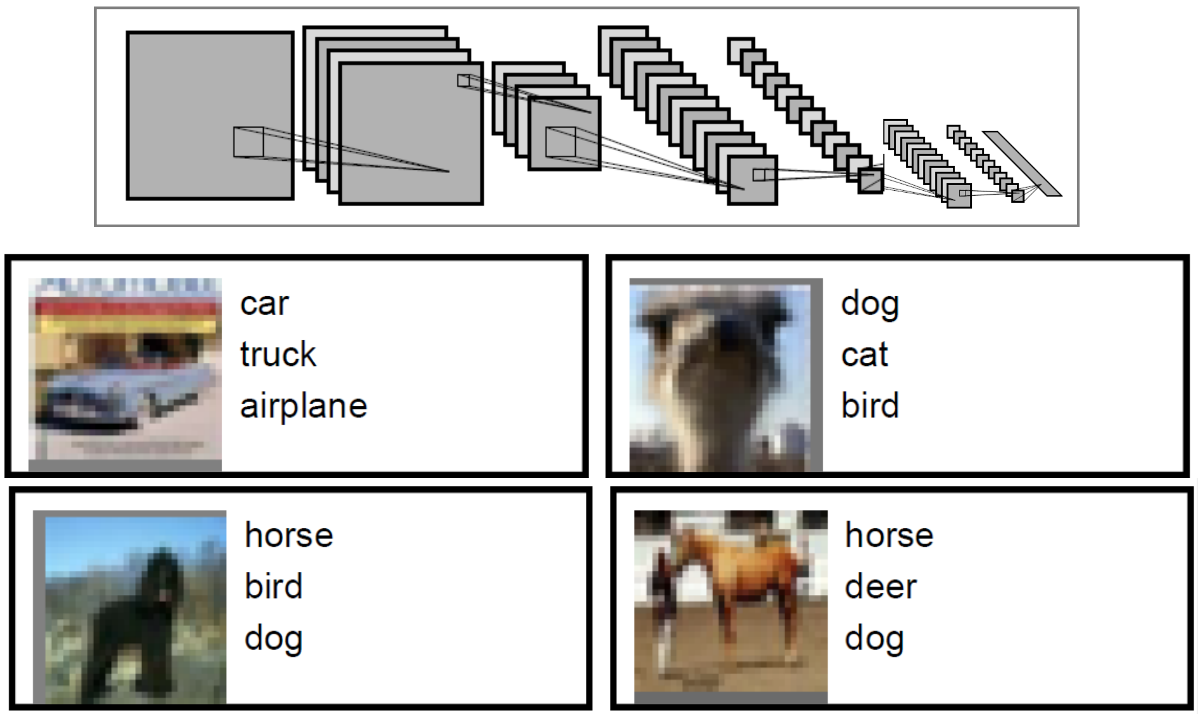
  Problem Set 1,
  due: Friday, Jun 4

Neural Networks

Model Evaluation

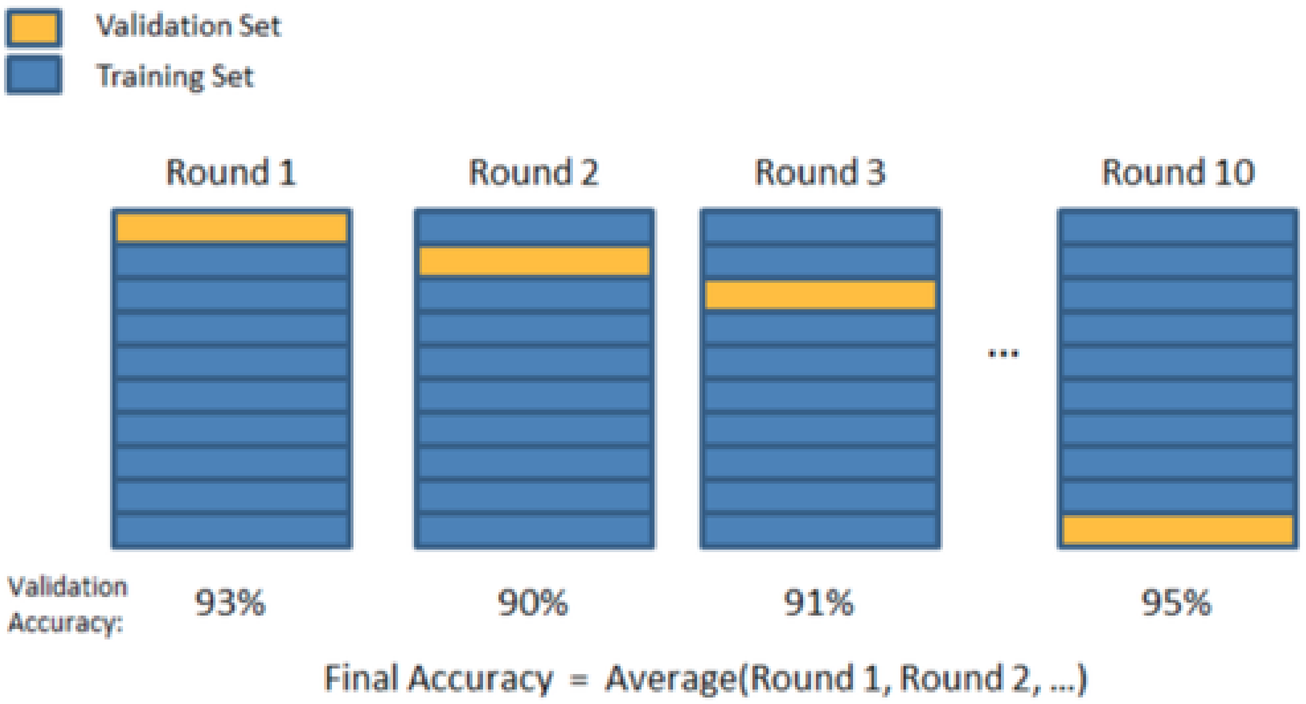
Top-k Accuracy

- Show top three most likely classes



N-Fold Cross Validation

- What if we don't have enough data for train/test/validation sets?
- Solution: use N-fold cross validation.
- Split training set into train/validation sets N times.
- Report average predictions over N val sets, e.g. N=10:



Confusion Matrix

- A performance measurement for machine learning classification problem where output can be two or more classes.

- **True Positive:**
  predicted positive and it’s true

  True Negative:
  predicted negative and it’s true

  False Positive:
  predicted positive and it’s false

  False Negative:
  predicted negative and it’s false

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

Recall

- Out of all the positive classes, how much we predicted correctly. It should be high as possible.

$$\textit{Recall} = \frac{TP}{TP + FN}$$

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

Precision

- Out of all the positive classes we have predicted, how many are actually positive.

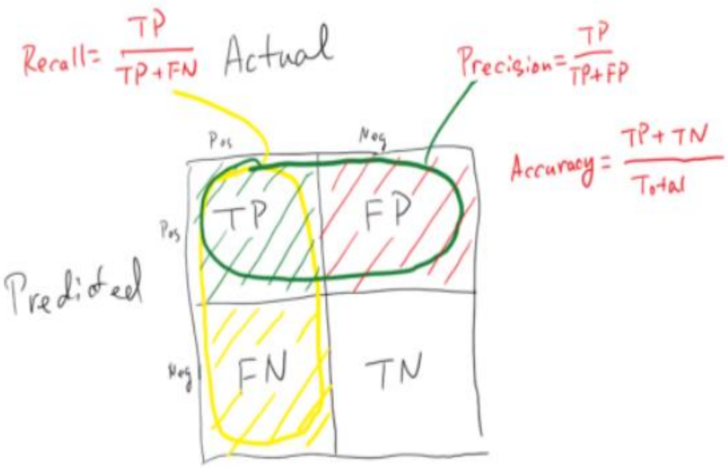
$$\textit{Precision} = \frac{TP}{TP + FP}$$

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

Example

y	y pred	output for threshold 0.6	Recall	Precision	Accuracy
0	0.5	0	**1/2**	**2/3**	**4/7**
1	0.9	1			
0	0.7	1			
1	0.7	1			
1	0.3	0			
0	0.4	0			
1	0.5	0			



F-measure

- F-score helps to measure Recall and Precision at the same time.

$$\textit{F - measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

Alex Krizhevsky

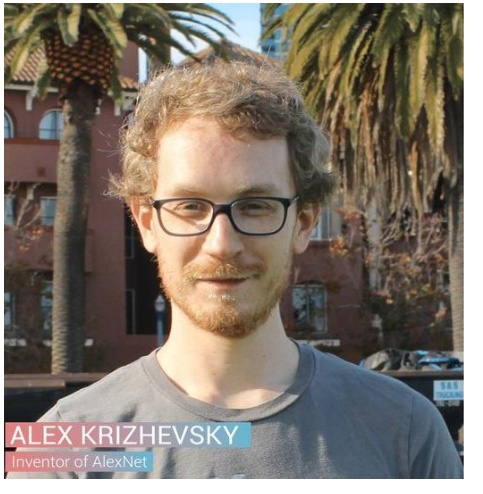
Alex Krizhevsky

FOLLOW

Dessa
Verified email at dessa.com
Machine Learning

TITLE	CITED BY	YEAR
Imagenet classification with deep convolutional neural networks A Krizhevsky, I Sutskever, GE Hinton Advances in neural information processing systems, 1097-1105	57520	2012
Dropout: a simple way to prevent neural networks from overfitting N Srivastava, G Hinton, A Krizhevsky, I Sutskever, R Salakhutdinov The journal of machine learning research 15 (1), 1929-1958	18013	2014

Hence the name ***AlexNet***



ACM Turing Award (2019)

- Three ‘Godfathers of Deep Learning’ Selected for Turing Award

- ***Geoff Hinton***, an emeritus professor at the University of Toronto and a senior researcher at Alphabet Inc.’s Google Brain
  https://arxiv.org/pdf/1805.09092.pdf

- ***Yann LeCun***, a professor at New York University and the chief AI scientist at Facebook Inc.

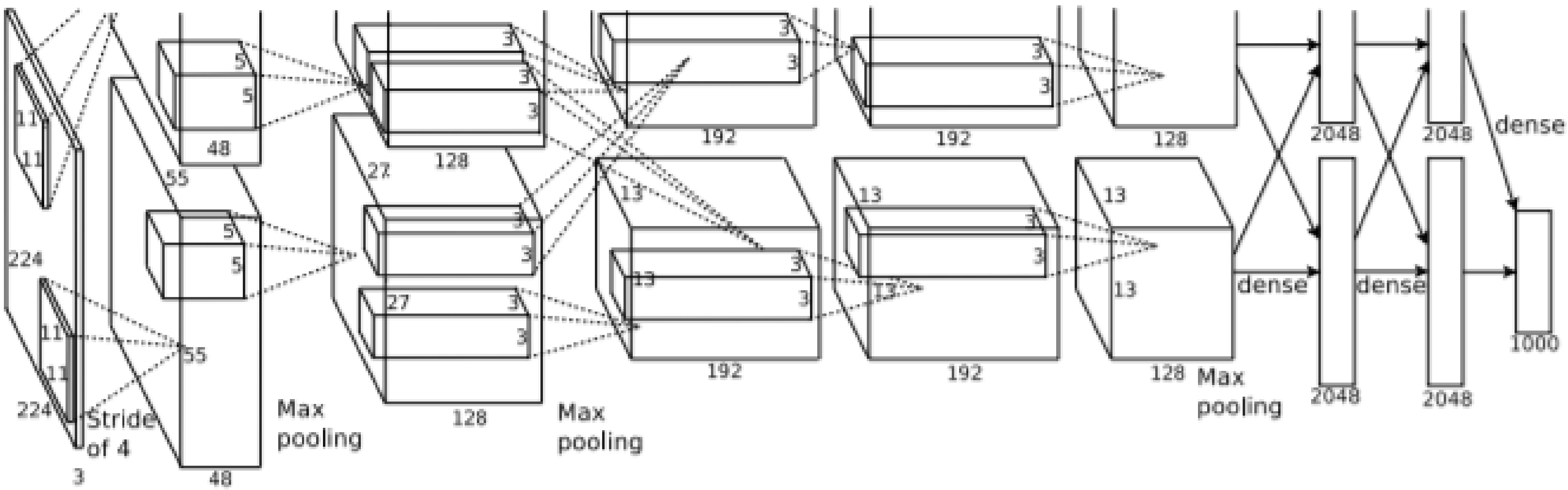
- ***Yoshua Bengio***, a professor at the University of Montreal as well as co-founder of AI company Element AI Inc.

Geoffrey E Hinton

Yann LeCun

Yoshua Bengio

The AlexNet


224
224
3
11
11
11
11
Stride of 4
5
5
5
5
55
55
48
48
Max pooling
27
27
128
128
3
3
3
3
3
3
3
3
Max pooling
13
13
192
192
3
3
3
3
13
13
192
192
3
3
3
3
13
13
128
128
Max pooling
dense
dense
dense
2048
2048
2048
2048
1000

Data Augmentation

- A technique that prevents overfitting.

- How?
  By artificially enlarging the dataset using label-preserving transformations.

- Examples:
  - generating image translations and horizontal reflections
  - altering the intensities of the RGB channels in training images: add perturbations to each RGB image pixel $I_{xy} = [I_{xy}^R, I_{xy}^G, I_{xy}^B]$

[Krizhevsky *et al.*]

Data Augmentation

- Could be computed "on the fly," and do not necessarily need to be stored on disk.

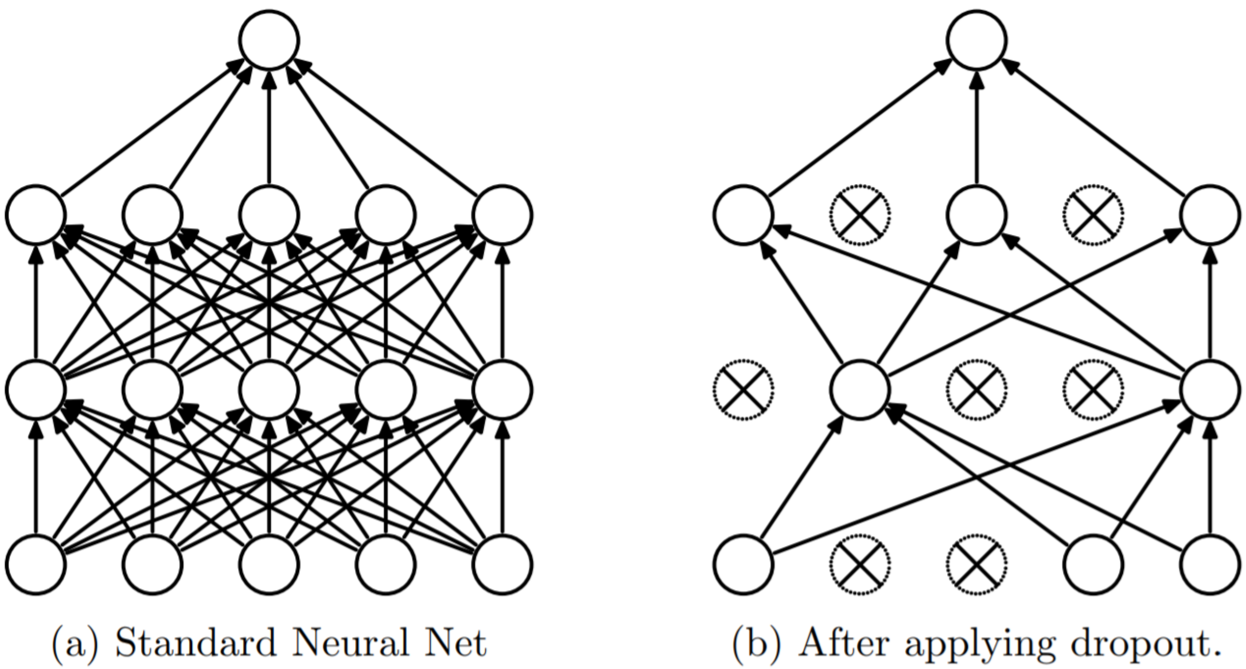
- How?
  The transformed images are generated in Python code on the CPU while the GPU is training on the previous batch of images.

- So these data augmentation schemes can be, in effect, computationally free.

[Krizhevsky *et al.*]

Dropout: A Classical Regularization Technique

- Many Deep Models employ dropout at training time to avoid overfitting, allowing for better generalization.

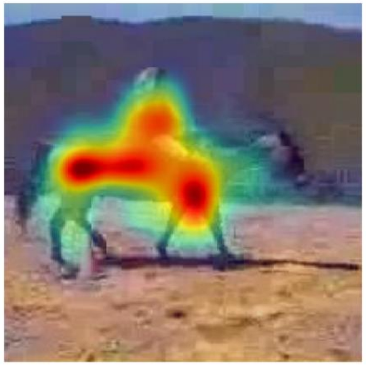


[Srivastava *et al.*]

Excitation Dropout

- We target answering the question: *Which neurons to drop out?*
  - *Neurons that have a higher contribution to the ground-truth prediction.*
  - *Example for ground-truth class HorseRiding:*

$image$ | $evidence$: p_{EB}



[Zunino A.*, Bargal S. A.*, Morerio P., Zhang J., Sclaroff S., & Murino V.
Excitation Dropout: Encouraging Plasticity in Deep Neural Networks, IJCV'21.]

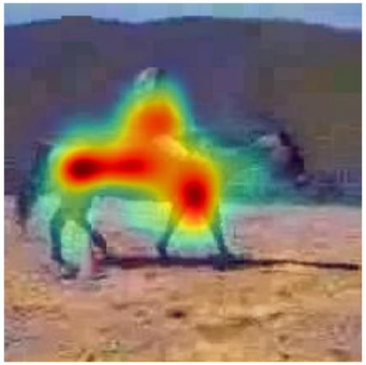
Excitation Dropout

- We define the retaining probability p for a neuron as follows:

$$p = 1 - \frac{(1-P)*(N-1)*\boxed{p_{EB}}}{((1-P)*N-1)*\boxed{p_{EB}}+P}$$

N: *# neurons*
P: *base probability*

$image$ $\qquad$ $evidence$: p_{EB}

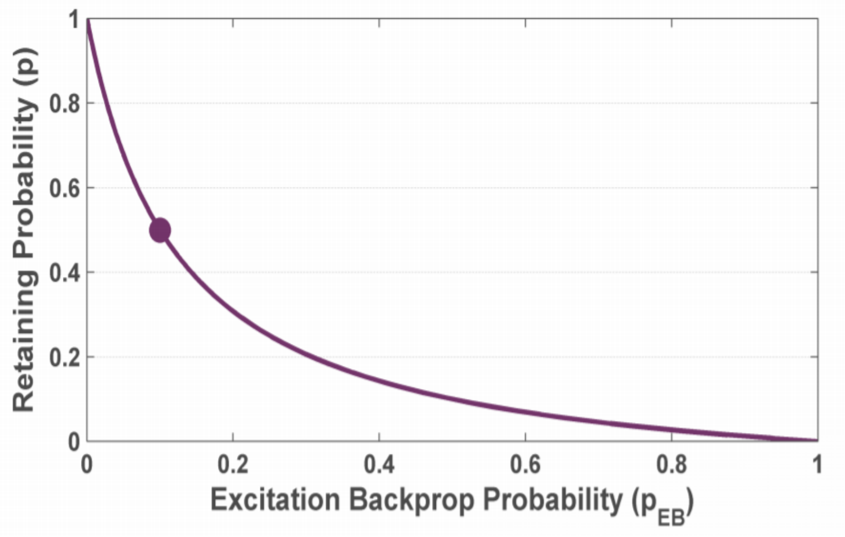


[Zunino A.*, Bargal S. A.*, Morerio P., Zhang J., Sclaroff S., & Murino V. *Excitation Dropout: Encouraging Plasticity in Deep Neural Networks*, IJCV'21.]

Excitation Dropout

- We define the retaining probability p for a neuron as follows:

$$p = 1 - \frac{(1-P) * (N-1) * p_{EB}}{((1-P) * N - 1) * p_{EB} + P}$$

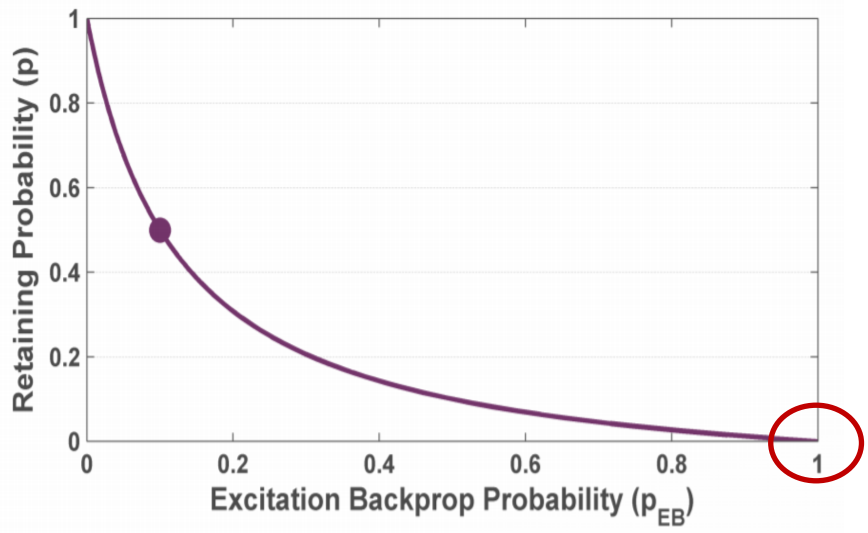


$N = 10, P = 0.5$

Excitation Dropout

- Constraint 1
  If a neuron has $p_{EB} = 1 \rightarrow p = 0$.

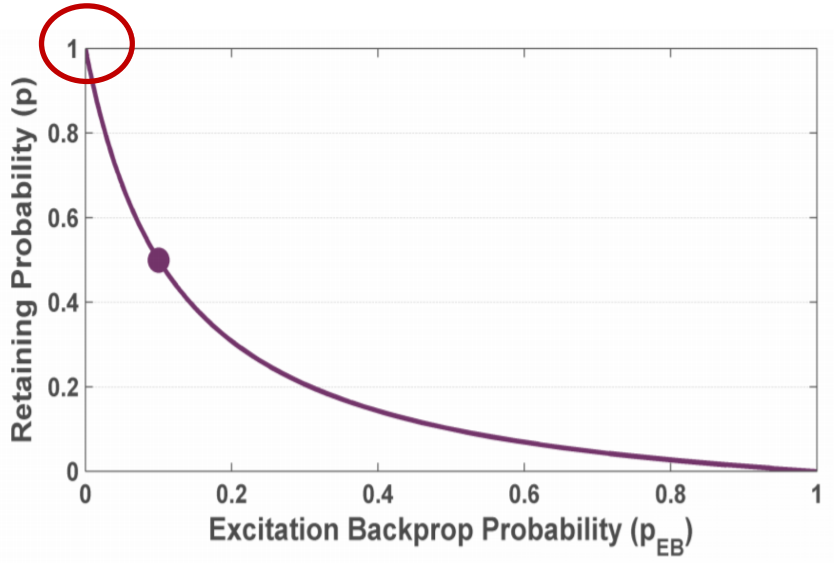
  "drop neurons which have a high contribution to the correct label"



Excitation Dropout

- Constraint 2
  If a neuron has $p_{EB} = 0 \rightarrow p = 1.$

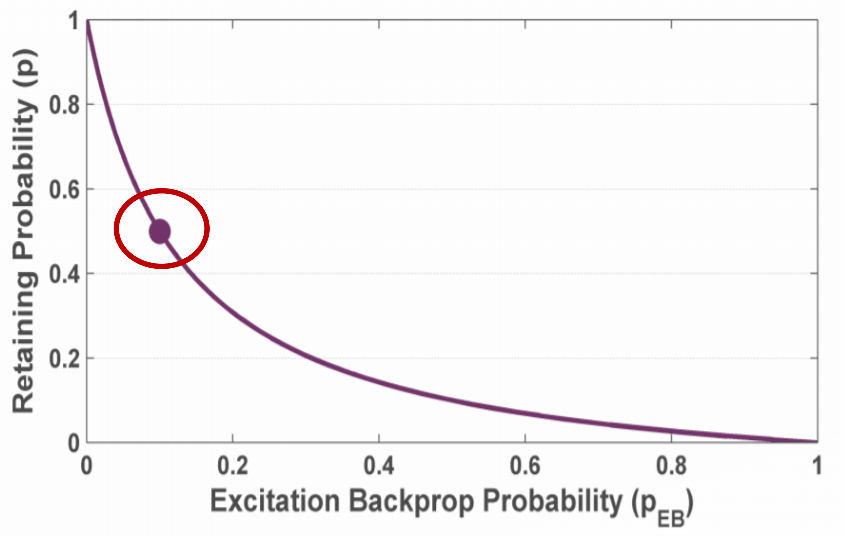
"keep neurons that do not contribute to the correct label"



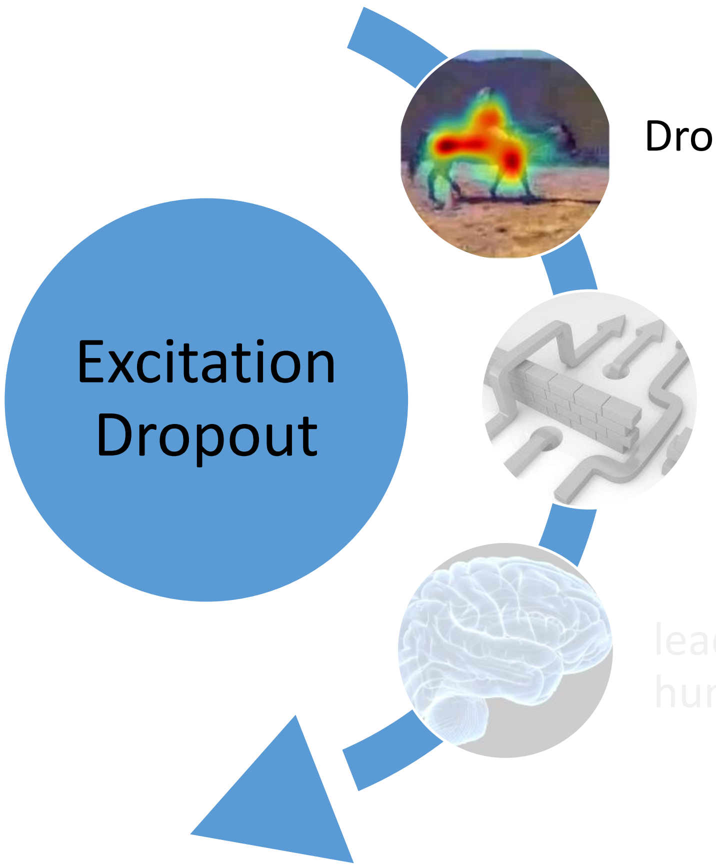
Excitation Dropout

- Constraint 3
  If a neuron has $p_{EB} = 1/N \rightarrow p = P$.

  "keep neurons with base probability P for their 'uniform' contribution"



Our Approach



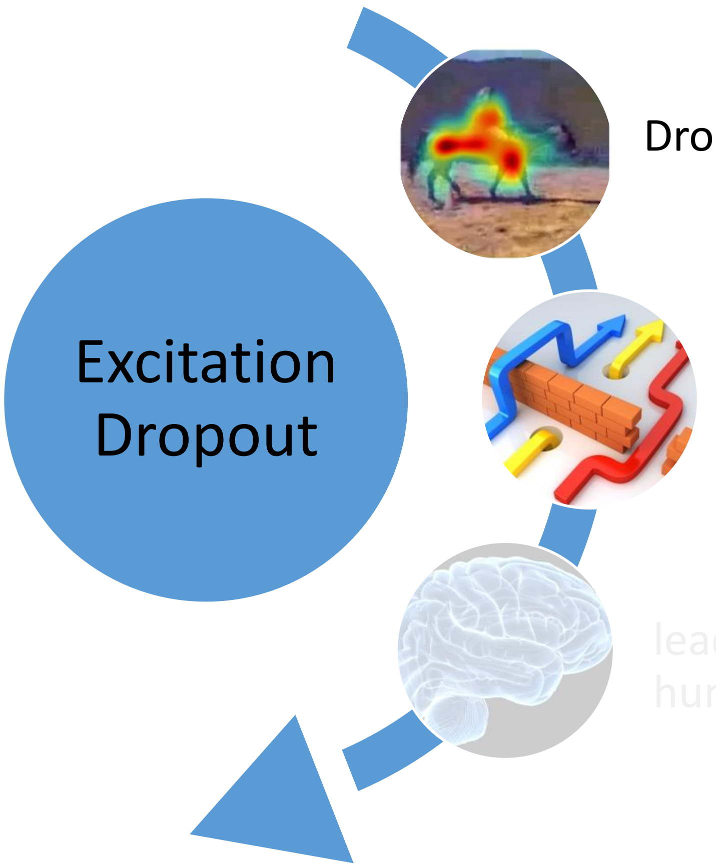
Dropping neurons having stronger evidence,

forces a network to learn alternative paths in order to maintain loss minimization,

leading to plasticity-like behavior, a characteristic of human brains too.

[Zunino A.*, Bargal S. A.*, Morerio P., Zhang J., Sclaroff S., & Murino V. *Excitation Dropout: Encouraging Plasticity in Deep Neural Networks*, IJCV'21.]

Our Approach



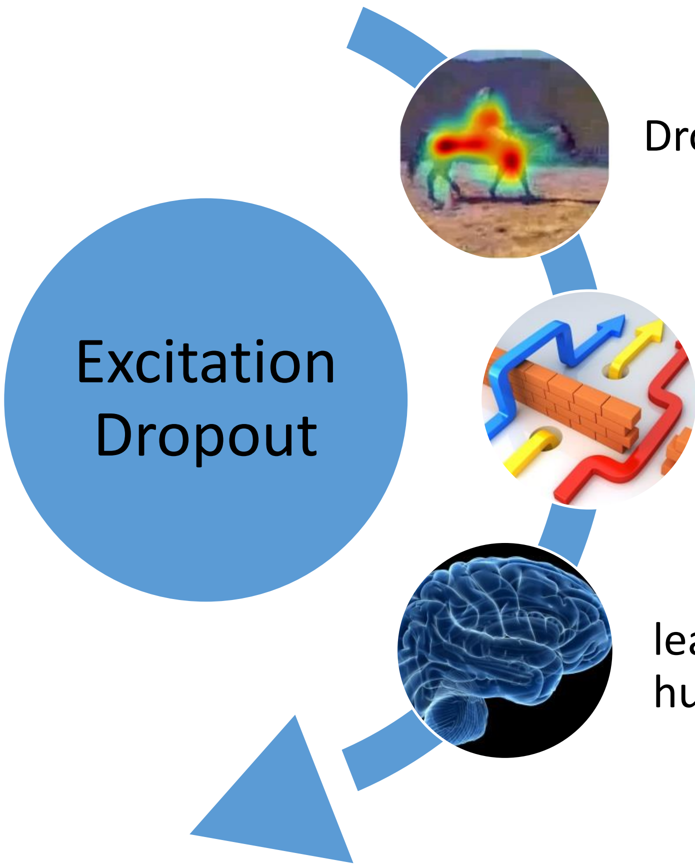
Dropping neurons having stronger evidence,

forces a network to learn alternative paths in order to maintain loss minimization,

leading to plasticity-like behavior, a characteristic of human brains too.

[Zunino A.*, Bargal S. A.*, Morerio P., Zhang J., Sclaroff S., & Murino V. *Excitation Dropout: Encouraging Plasticity in Deep Neural Networks*, IJCV'21.]

Our Approach



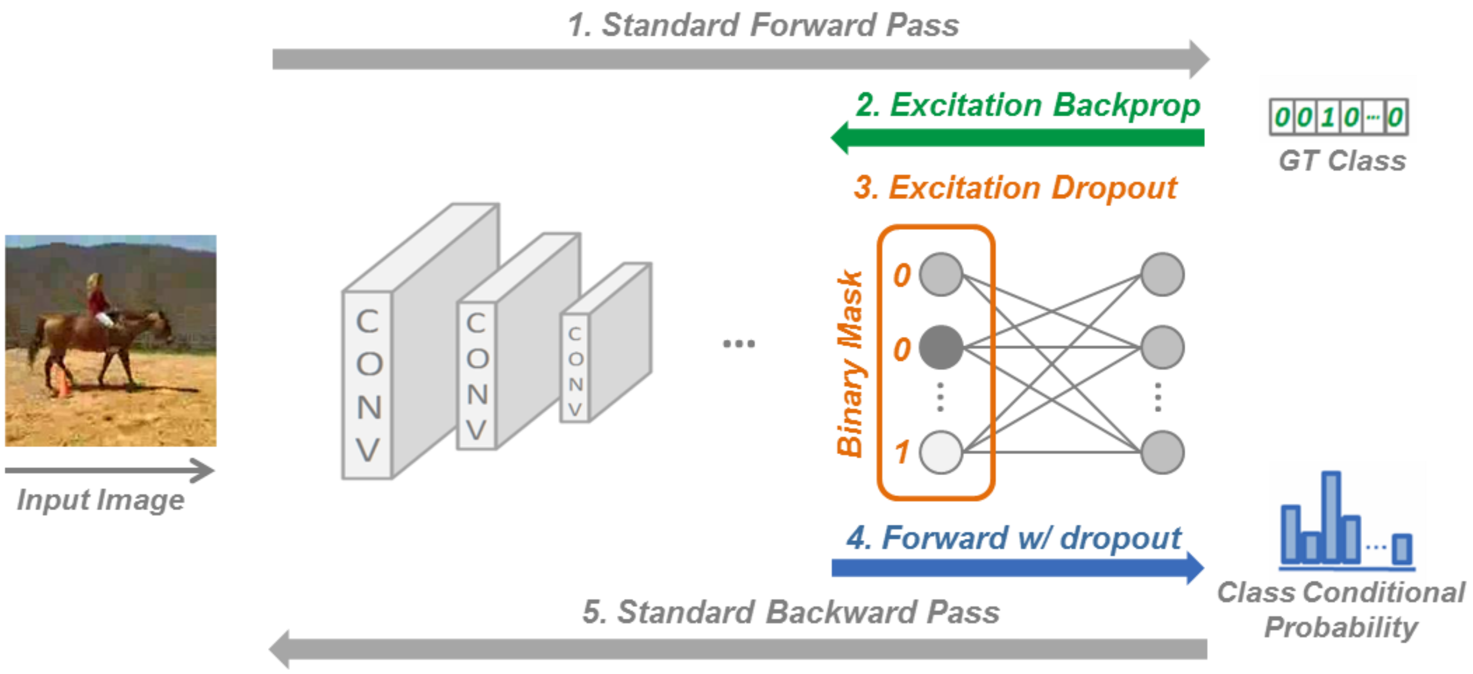
Dropping neurons having stronger evidence,

forces a network to learn alternative paths in order to maintain loss minimization,

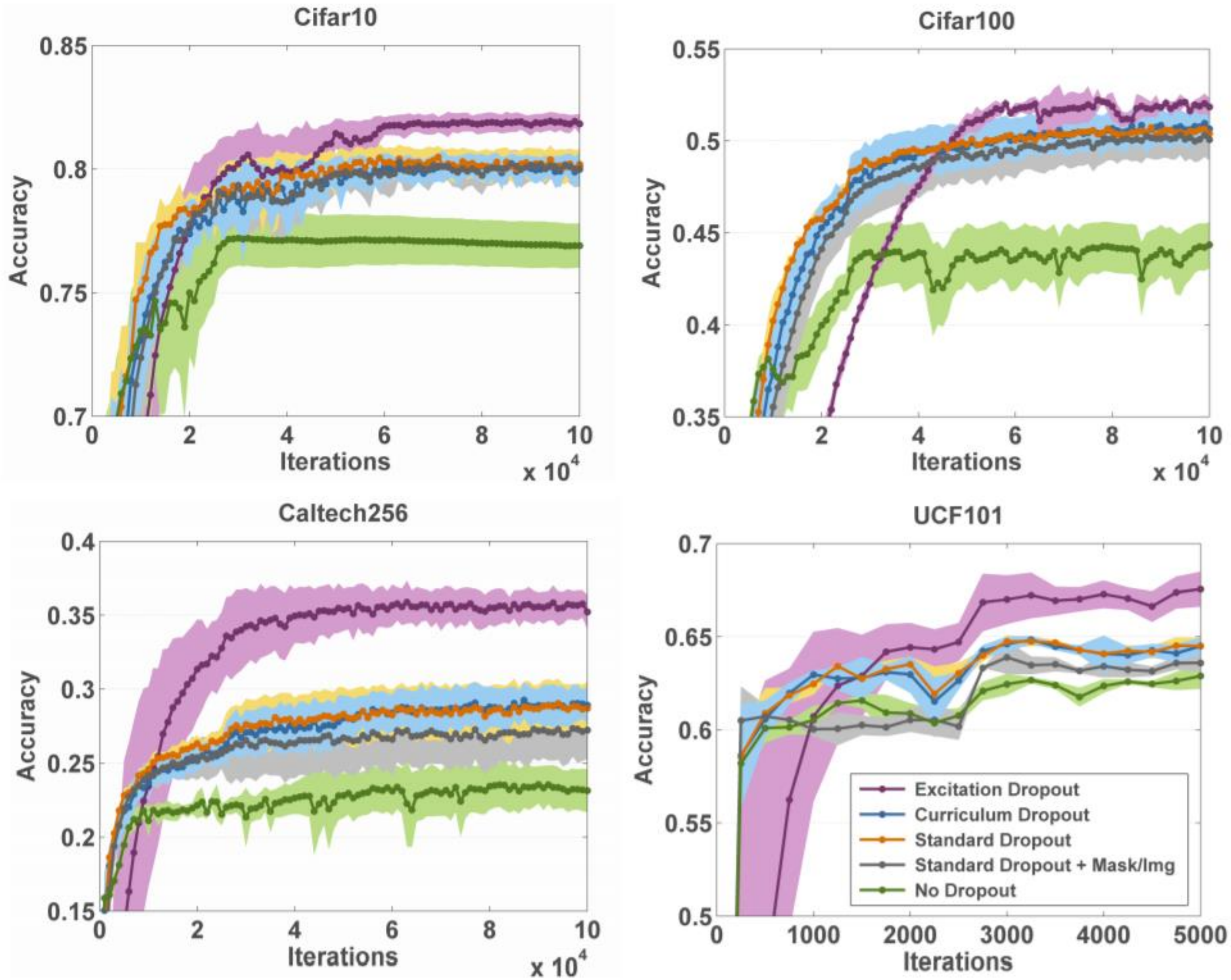
leading to plasticity-like behavior, a characteristic of human brains too.

[Zunino A.*, Bargal S. A.*, Morerio P., Zhang J., Sclaroff S., & Murino V. *Excitation Dropout: Encouraging Plasticity in Deep Neural Networks*, IJCV'21.]

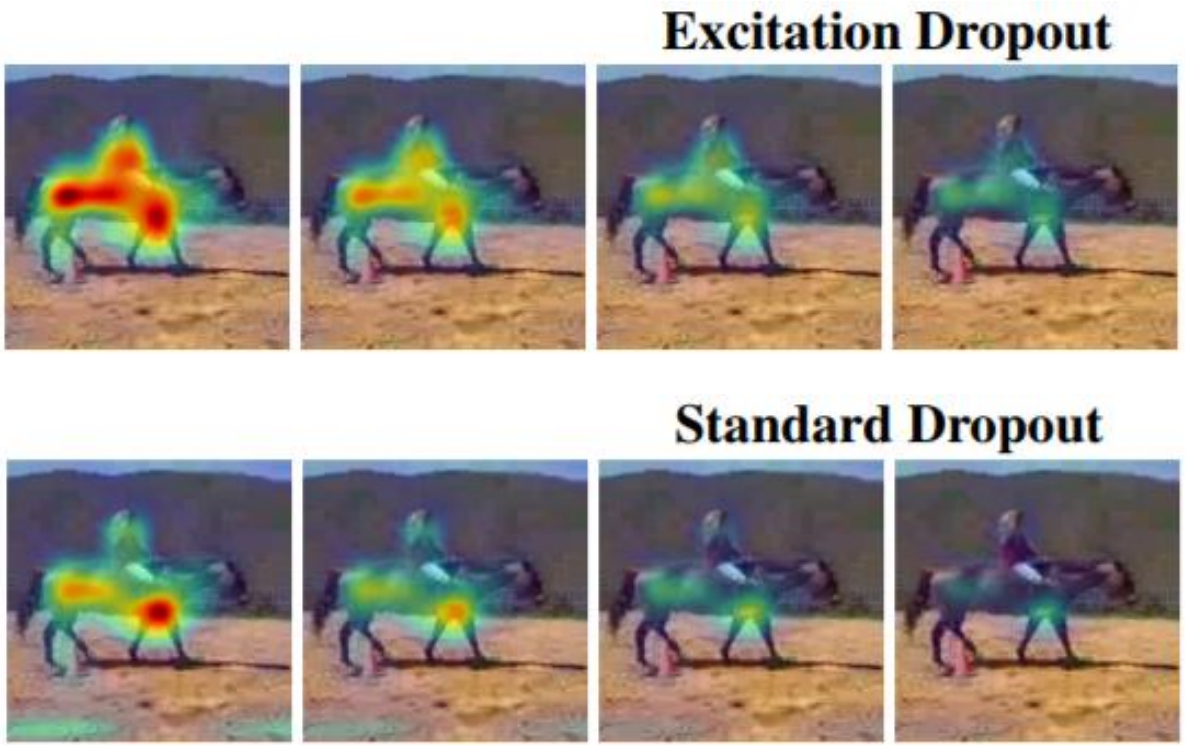
Excitation Dropout Pipeline



Improved Generalization



Robustness: Sample Result for UCF101

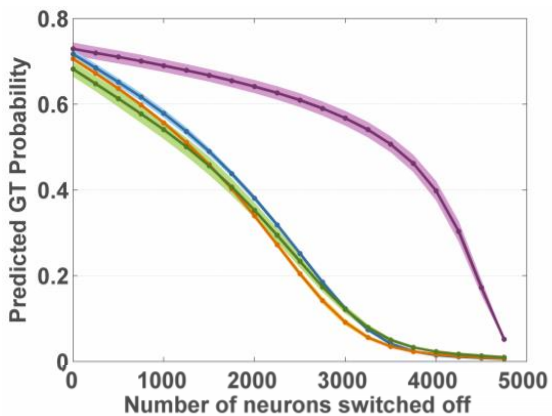


More neurons are dropped-out

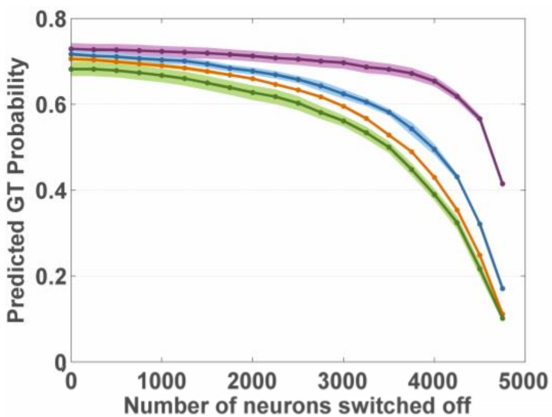
Robustness: Aggregate Results for UCF101

- Higher resilience to network pruning:

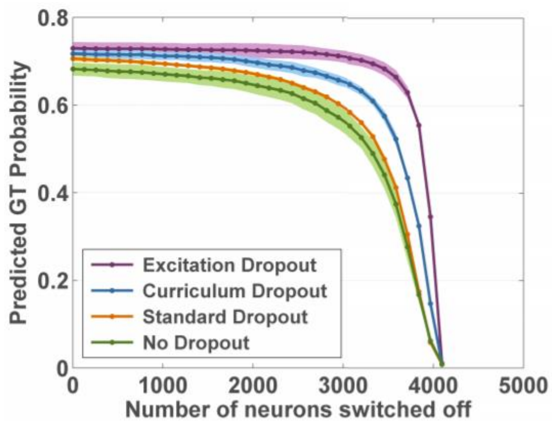
Most relevant



Least Relevant



Random



Conclusion: Excitation Dropout

- We propose a new regularization scheme that encourages the learning of alternative paths.

- We demonstrate that our approach yields:
  - better generalization on unseen data
  - higher utilization of network neurons
  - higher resilience to network pruning

Neural Networks

Explainability

Importance of *Explainability*

- An important action to be detected in the vision systems of autonomous vehicles is: *Pedestrian Crossing*

Importance of *Explainability*

Sample Misclassification

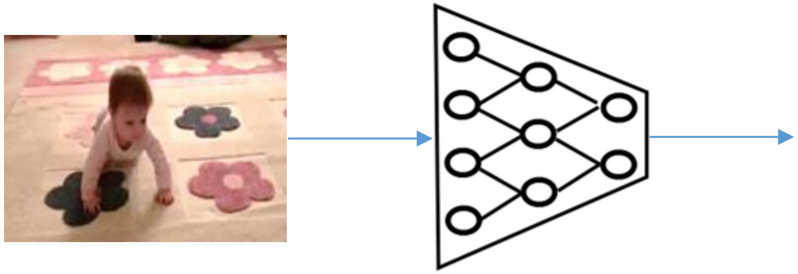
Ground Truth:
BabyCrawling

Classified as:
Pushups

Identifying Regions Responsible for a Prediction

- *How do we determine regions of an image that are responsible for a prediction?*

Vision CNN



prediction

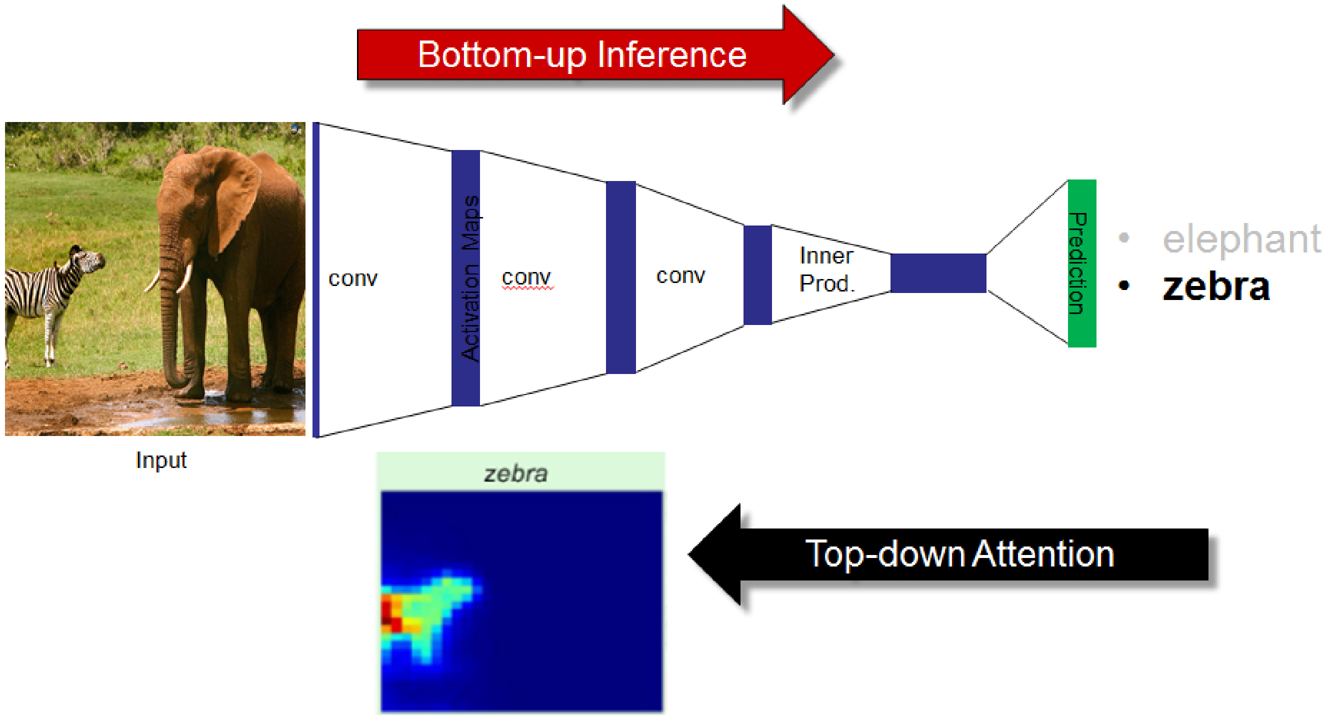
Black-box Methods

- *Input occlusion* — cover a part of the input image and see which part affect the classification the most.

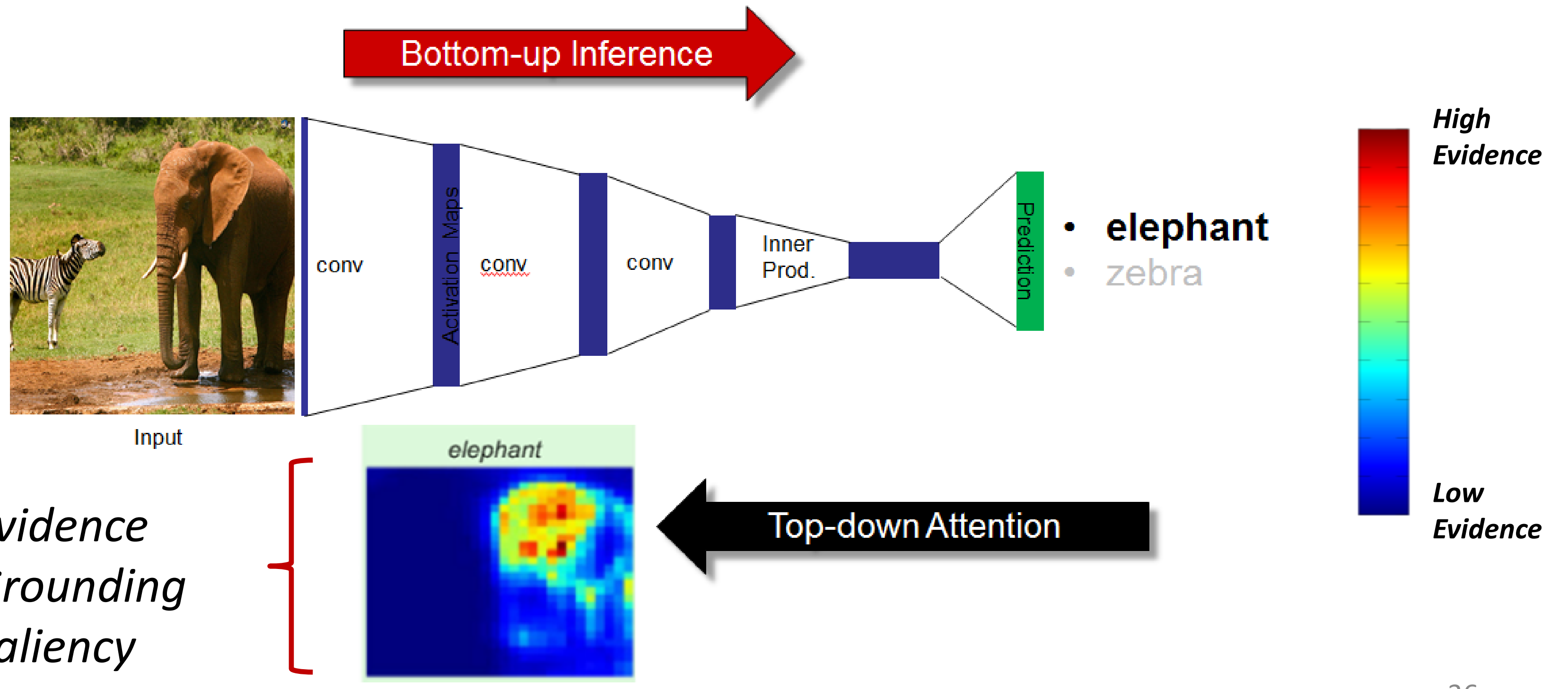
- For instance, given a trained image classification model, give the images below as input.

- If, for instance, we see that the 3rd image is classified with 98% probability as a dog, while the 2nd image only with 65% accuracy, it means that the eyes contribute significantly to the prediction.

White-box Methods



White-box Methods

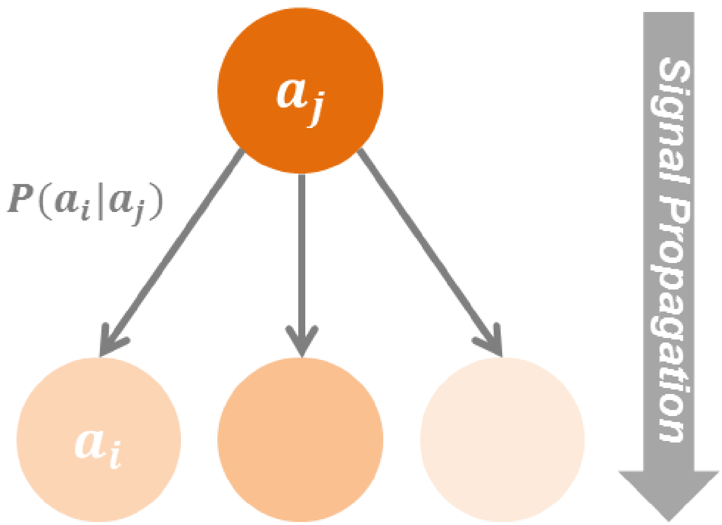


- *Evidence*
- *Grounding*
- *Saliency*

Excitation Backprop (EB)

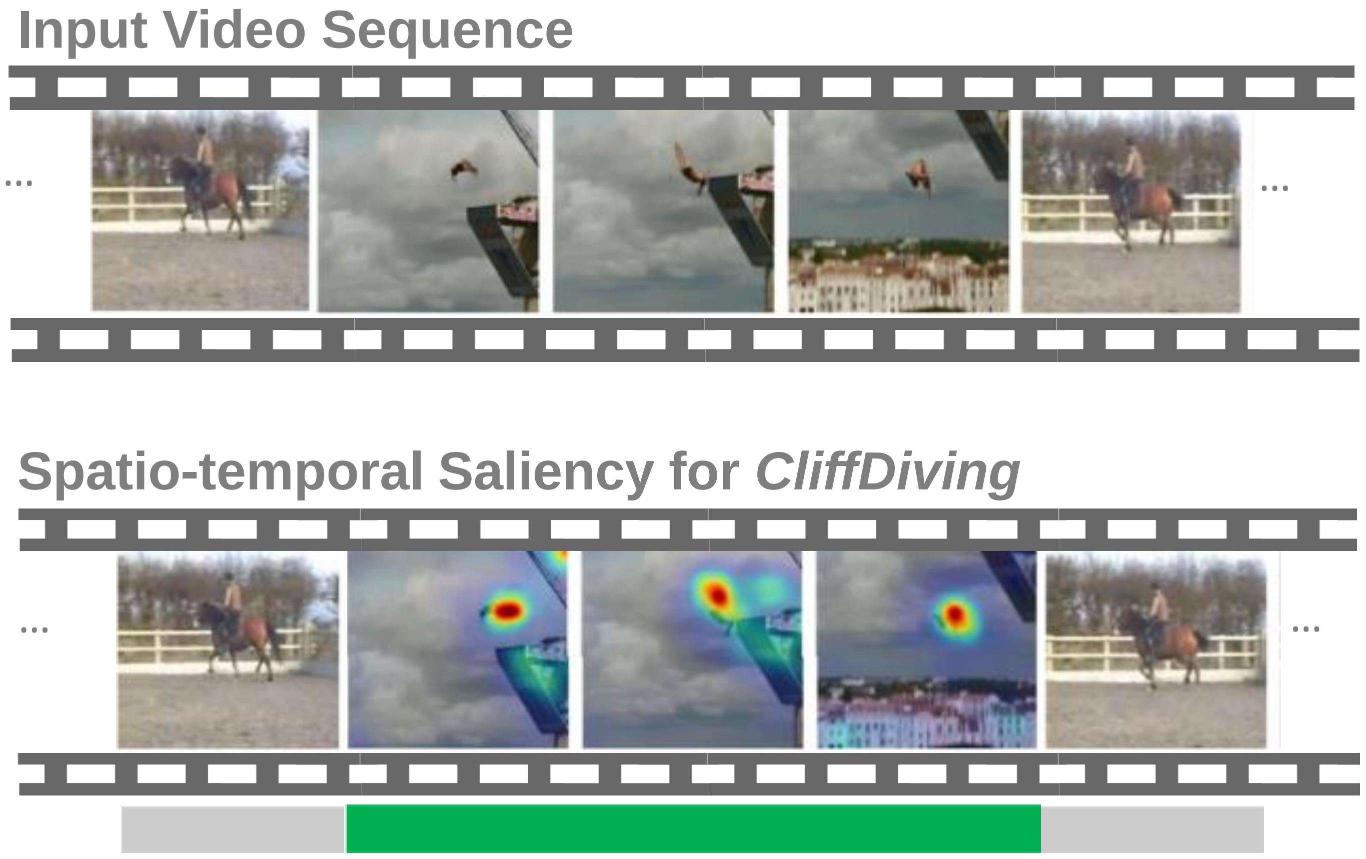
Only computed for excitatory connections $P(a_i|a_j)$

Value in Saliency Map $P(a_i) = \sum_{a_j \in \mathcal{P}_i} P(a_i|a_j)\, P(a_j)$



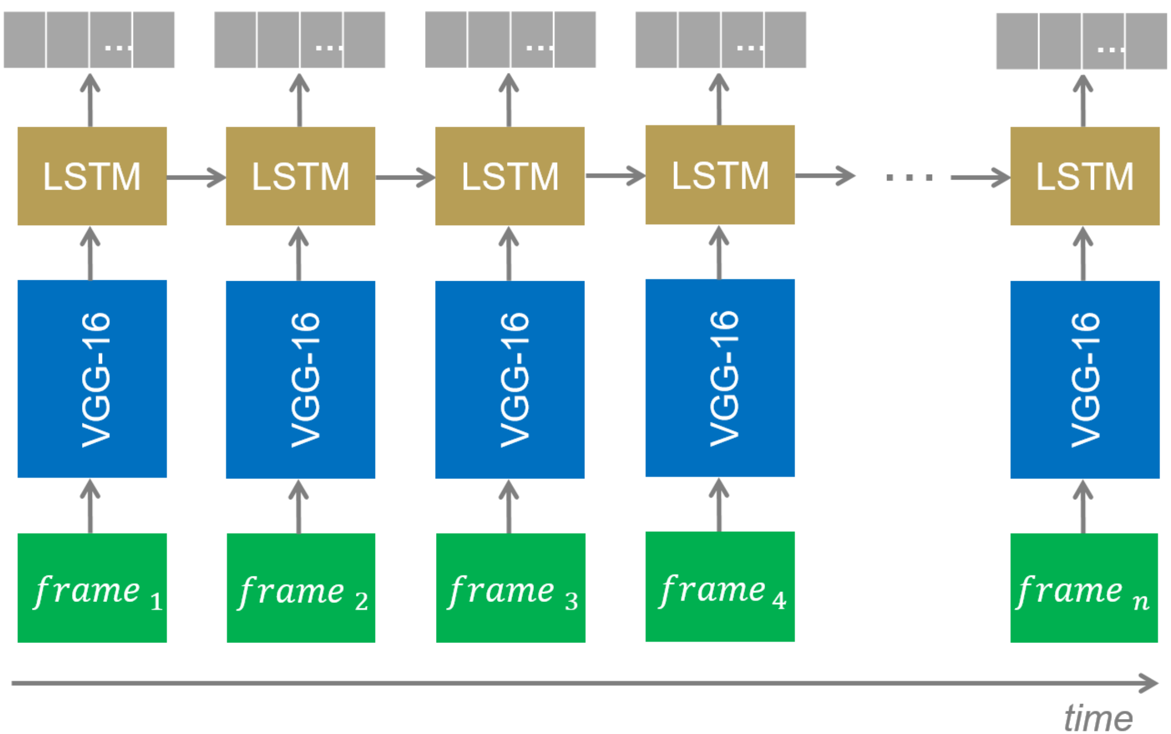
[Jianming Zhang, Zhe Lin, Jonathan Brandt, Xiaohui Shen, Stan Sclaroff. "*Top-down Neural Attention by Excitation Backprop., ECCV'16*]

Spatiotemporal Grounding

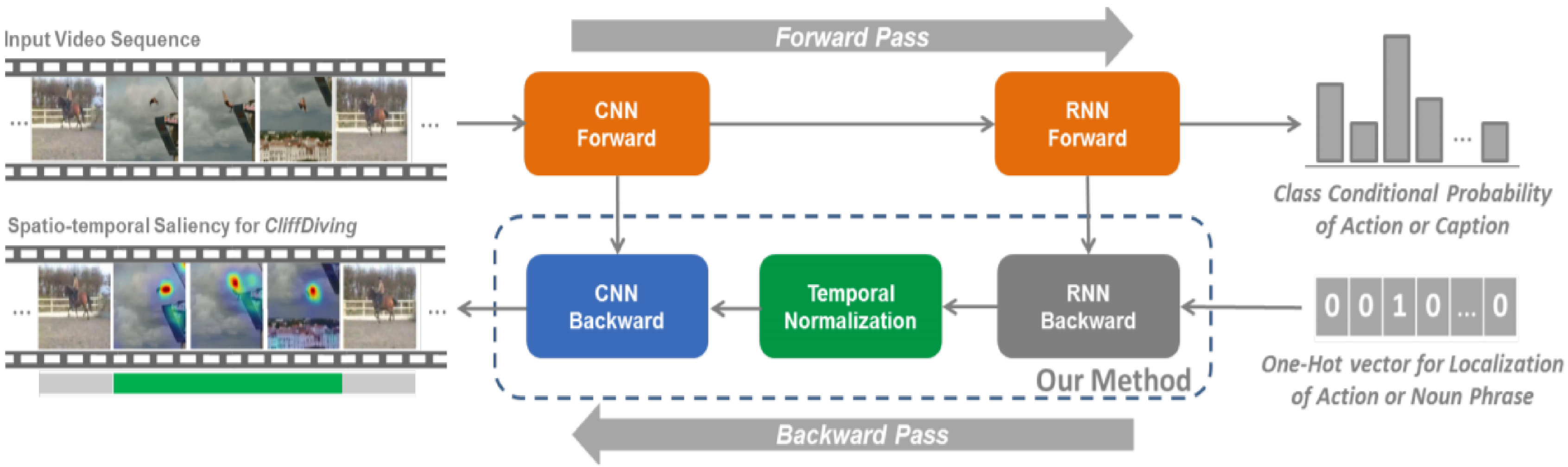


Architecture: Forward Pass

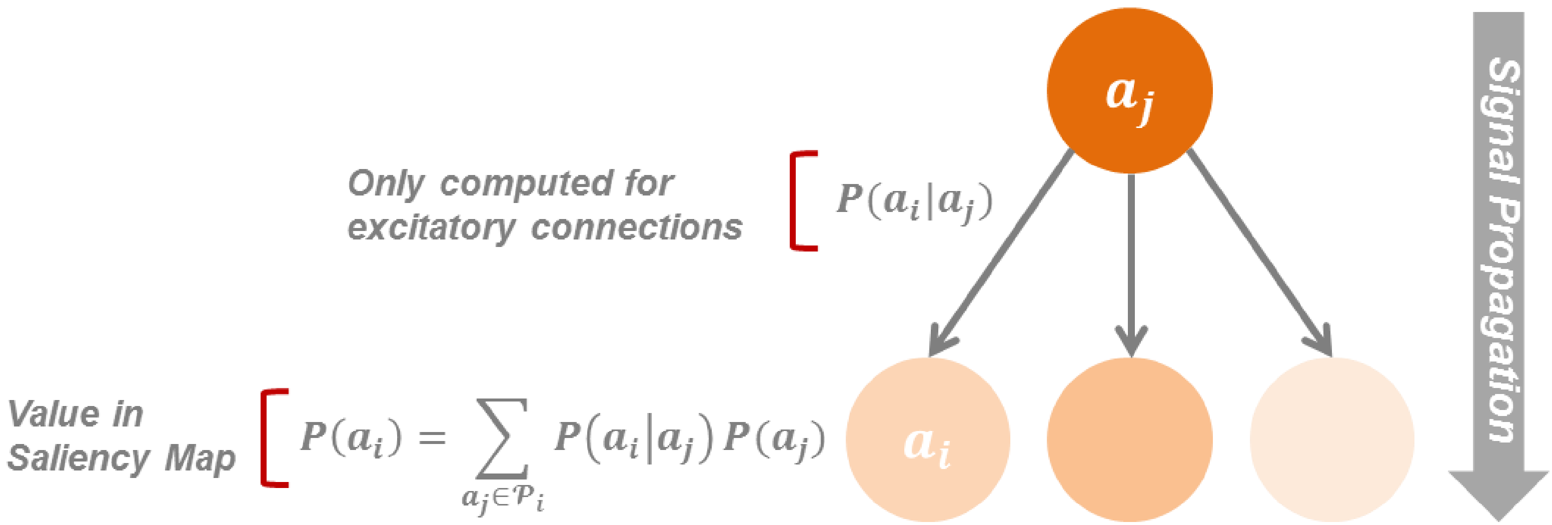
- CNN-LSTM is trained for the action recognition task.
- Resulting grounding is weakly-supervised.



Excitation Backprop in RNNs

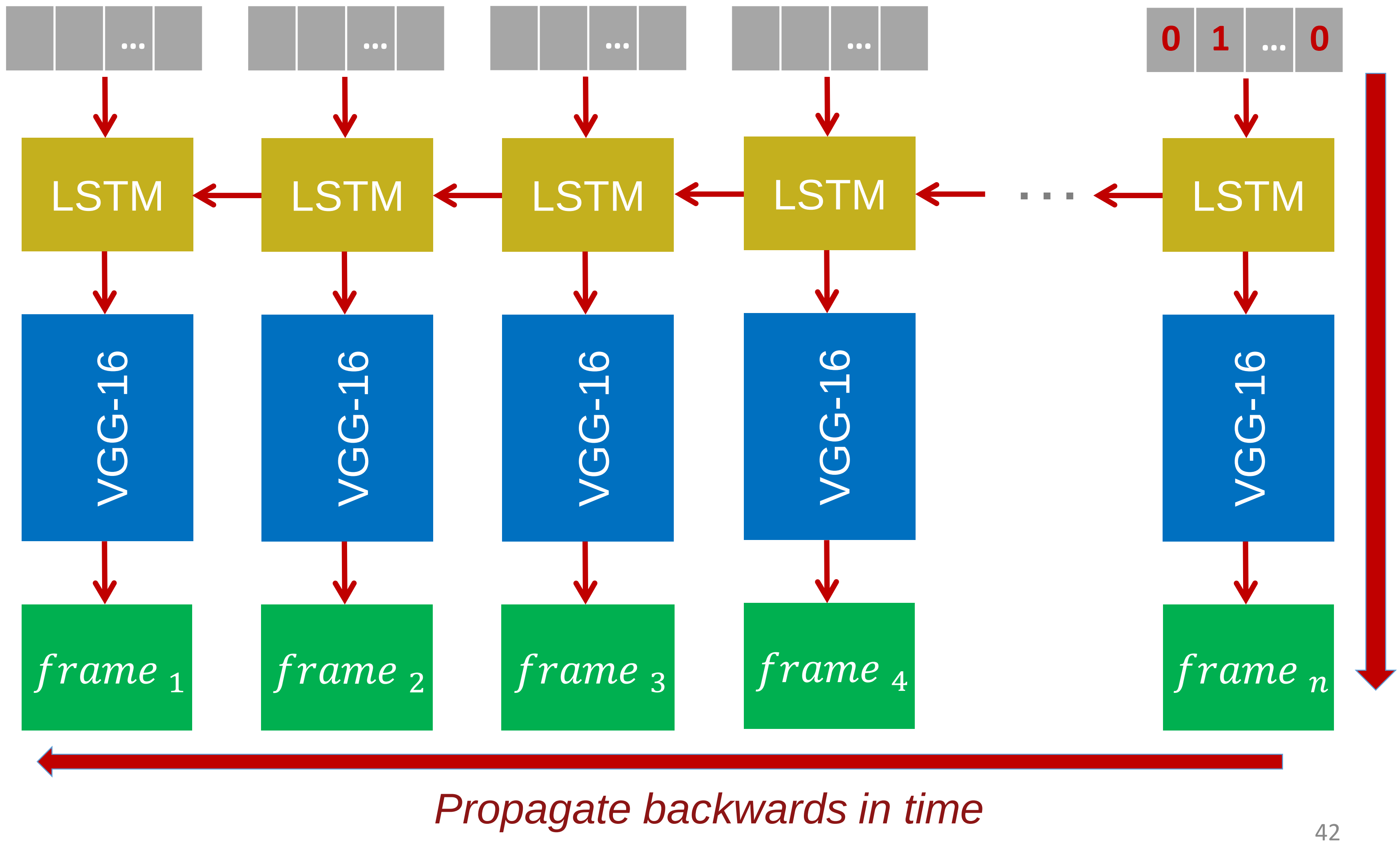


Excitation Backprop (EB)



[Jianming Zhang, Zhe Lin, Jonathan Brandt, Xiaohui Shen, Stan Sclaroff. "*Top-down Neural Attention by Excitation Backprop., ECCV'16*]

Architecture: Backward Grounding Pass



Applications

- Action Detection (*videos*)

- Caption Grounding (*images, videos*)

- Reflecting the Abstraction Capability of Models

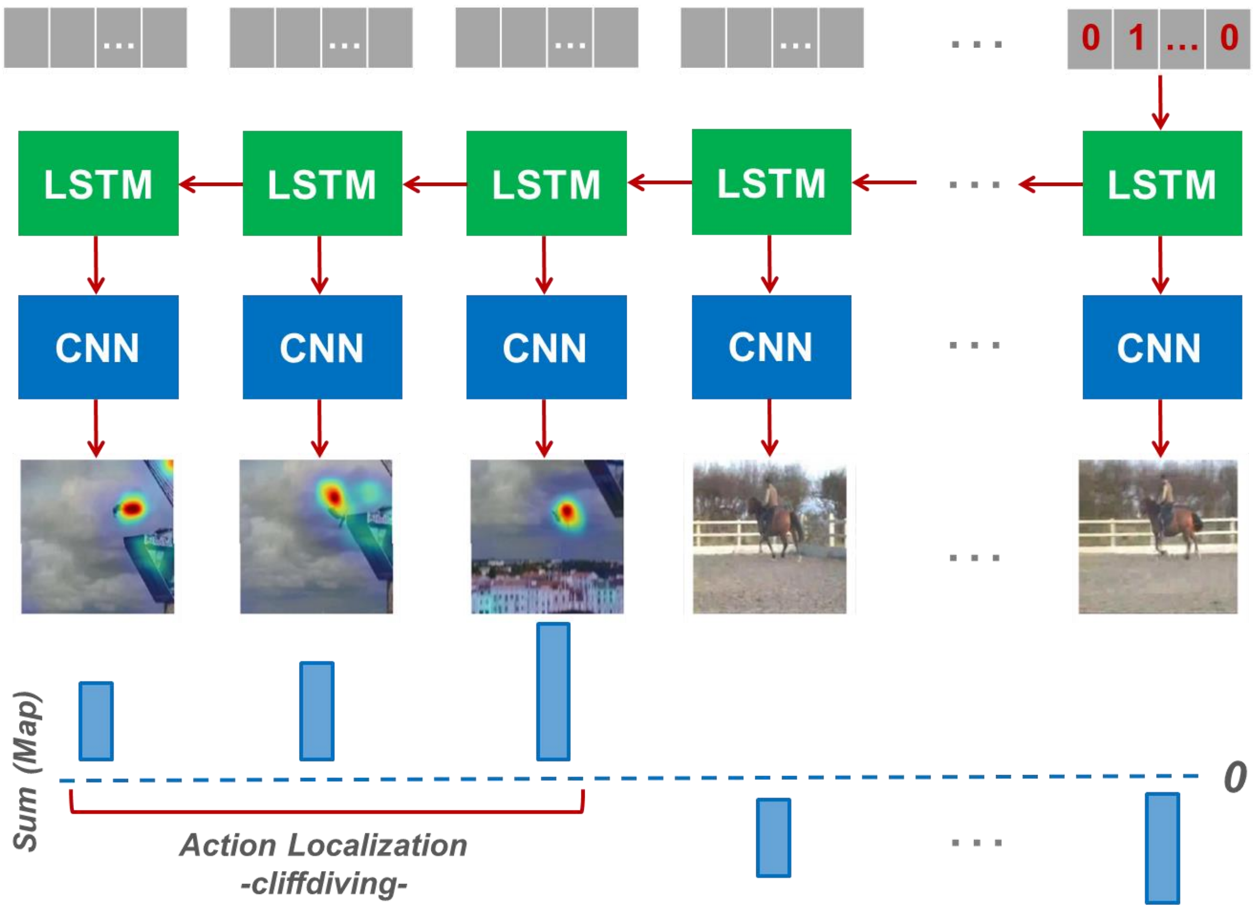
Applications

- Action Detection (*videos*)

- Caption Grounding (*images, videos*)

- Reflecting the Abstraction Capability of Models

Spatiotemporal Action Detection



Applications

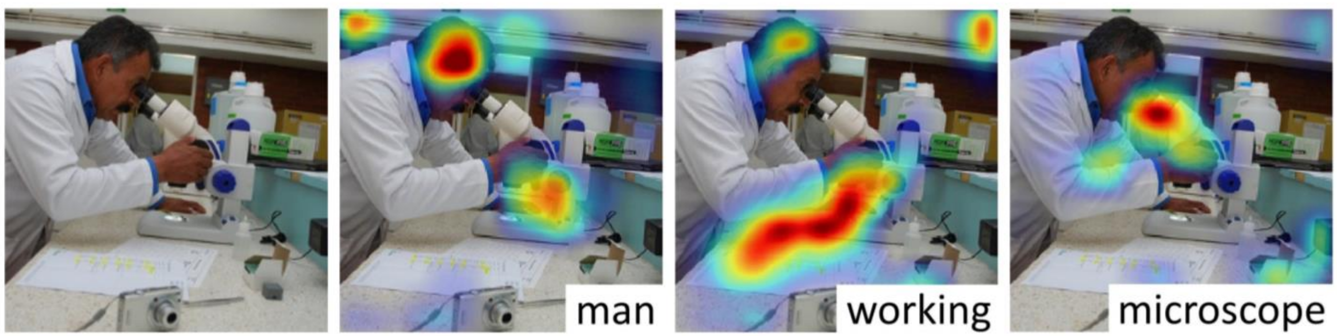
- Action Detection (*videos*)
- Caption Grounding (*images, videos*)
- Reflecting the Abstraction Capability of Models

Flicker30kEntities Dataset: Grounding Words of an Image Caption

image caption: *A man in a lab coat is working on a microscope.*

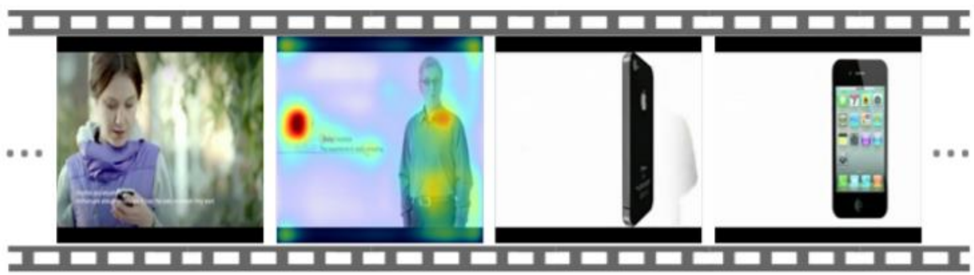
Flicker30kEntities Dataset: Grounding Words of an Image Caption

image caption: *A man in a lab coat is working on a microscope.*



MSRVTT Dataset: Grounding Words of a Video Caption

video caption: *“A man is talking about a phone”*



(a) grounding of the word *man*

(b) grounding of the word *phone*

Applications

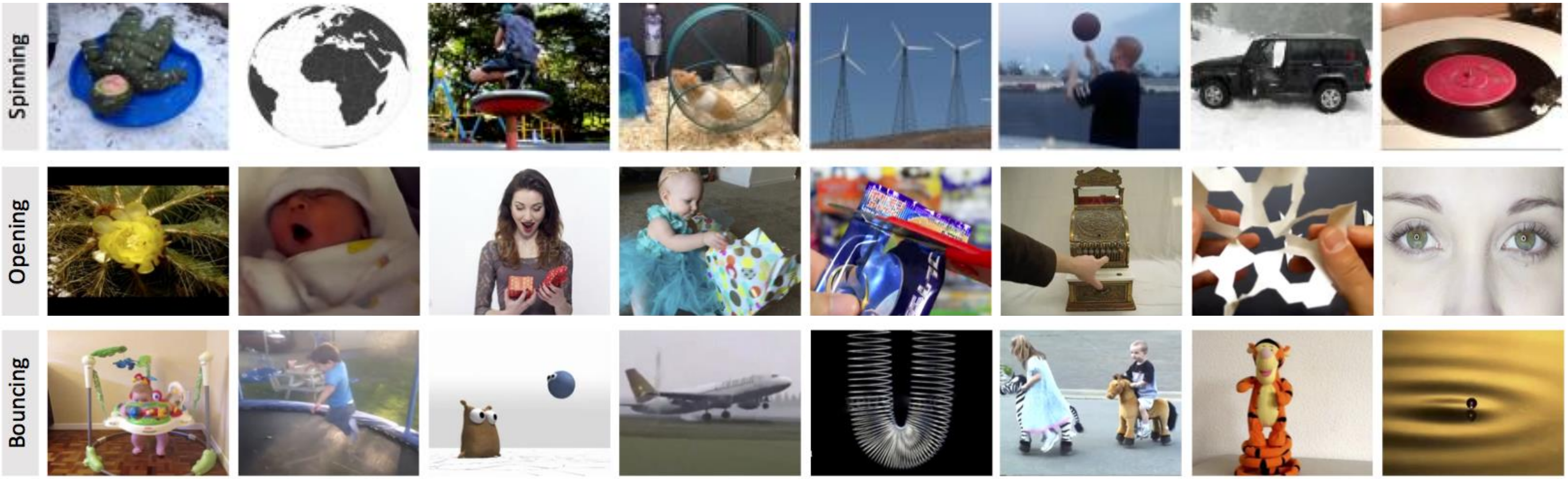
- Action Detection (*videos*)
- Caption Grounding (*images, videos*)
- Reflecting the Abstraction Capability of Models

Reflecting the Abstraction Capability of Models

- Moments in Time Dataset

  M. Monfort, A. Andonian, B. Zhou, K. Ramakrishnan, S. A. Bargal, T. Yan, L. Brown, Q. Fan, D. Gutfruend, C. Vondrick, A. Oliva. “Moments in Time Dataset: one million videos for event understanding.” *TPAMI*, 2019.

- Videos of abstract dynamical events performed by various actors.



Moments in Time Dataset

- Typically, classification accuracy is reported to summarize the recognition capability of models.

- However, classification accuracy alone is not representative as to whether the models are really modeling this diversity of actors.

- A classifier may be incorrectly classifying a whole subset of cases/actors.

Moments in Time Dataset

- Class: *Opening*