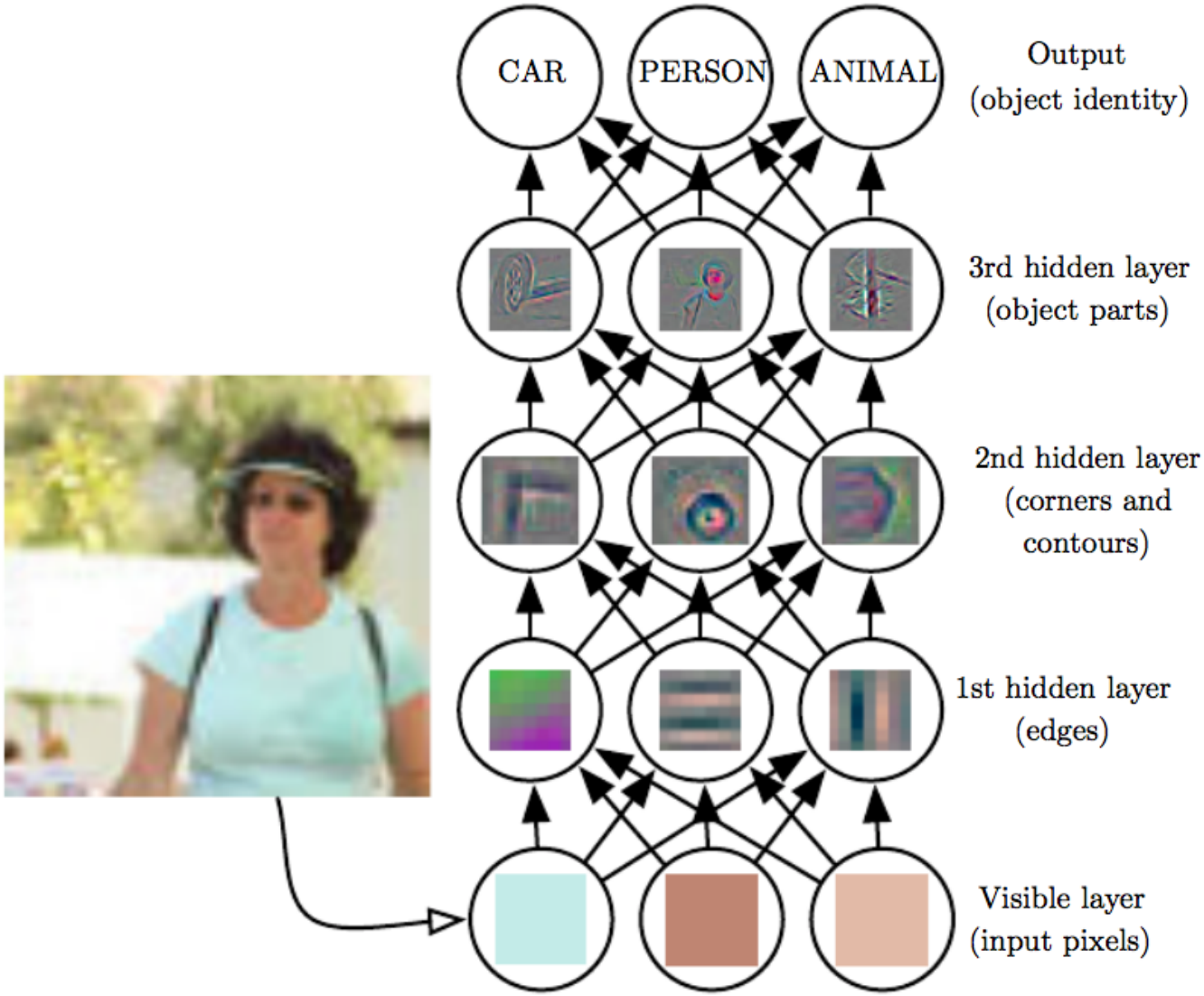
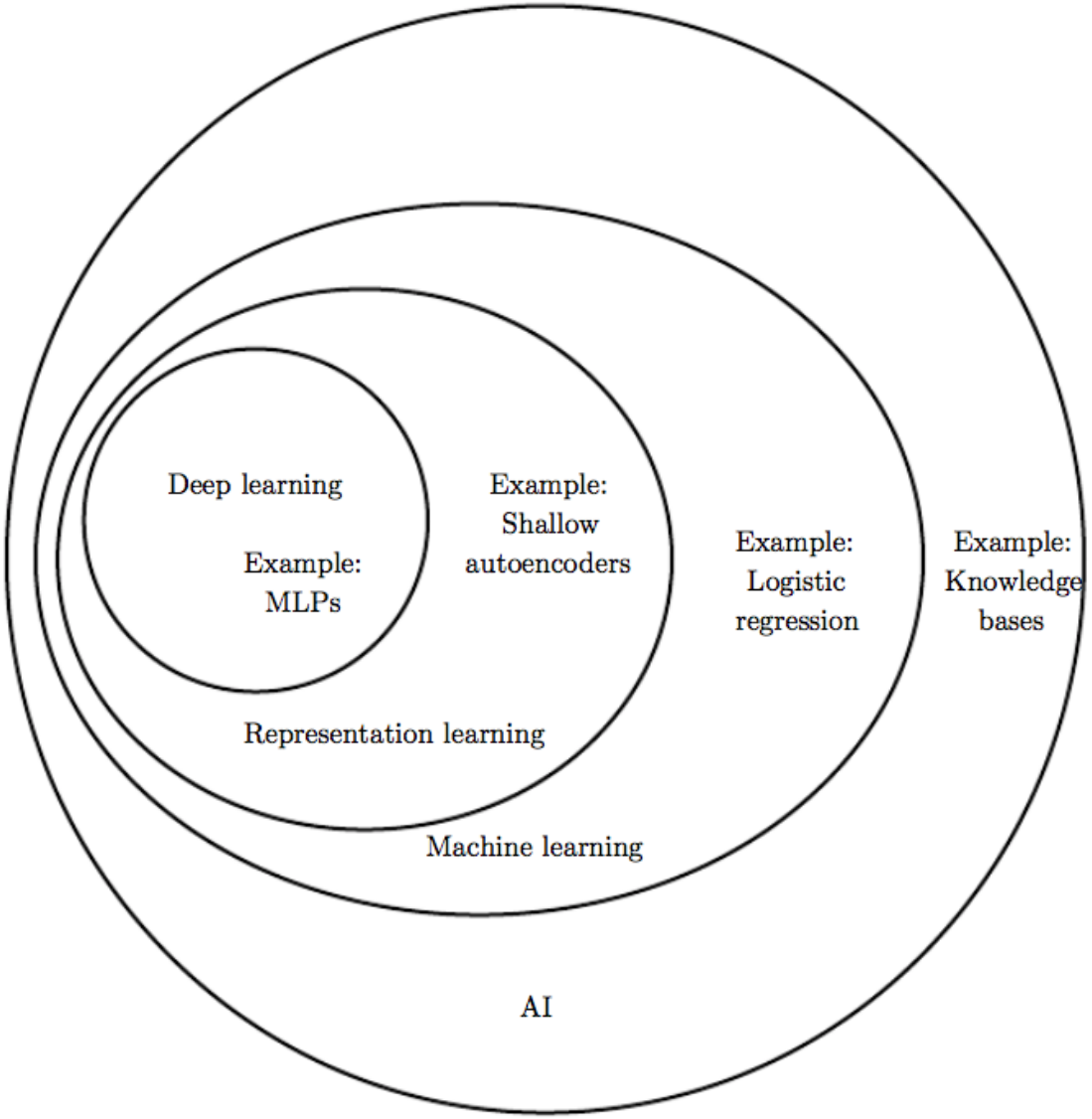
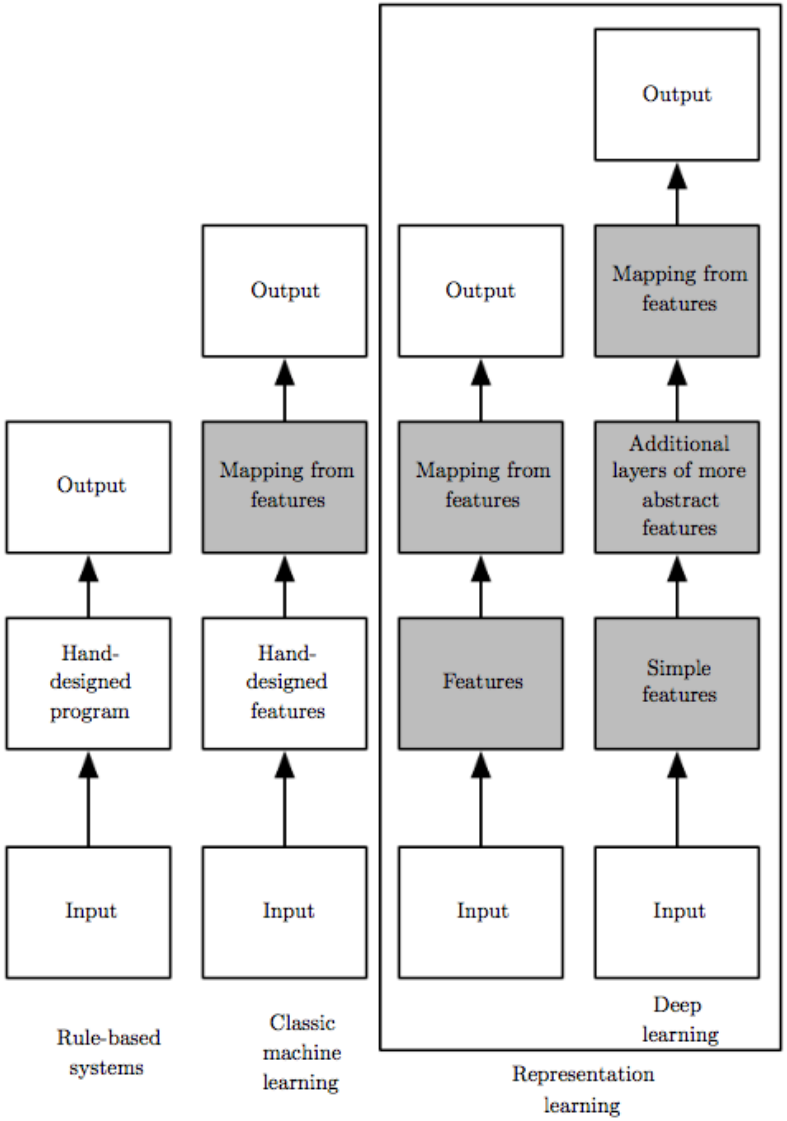
Deep Learning

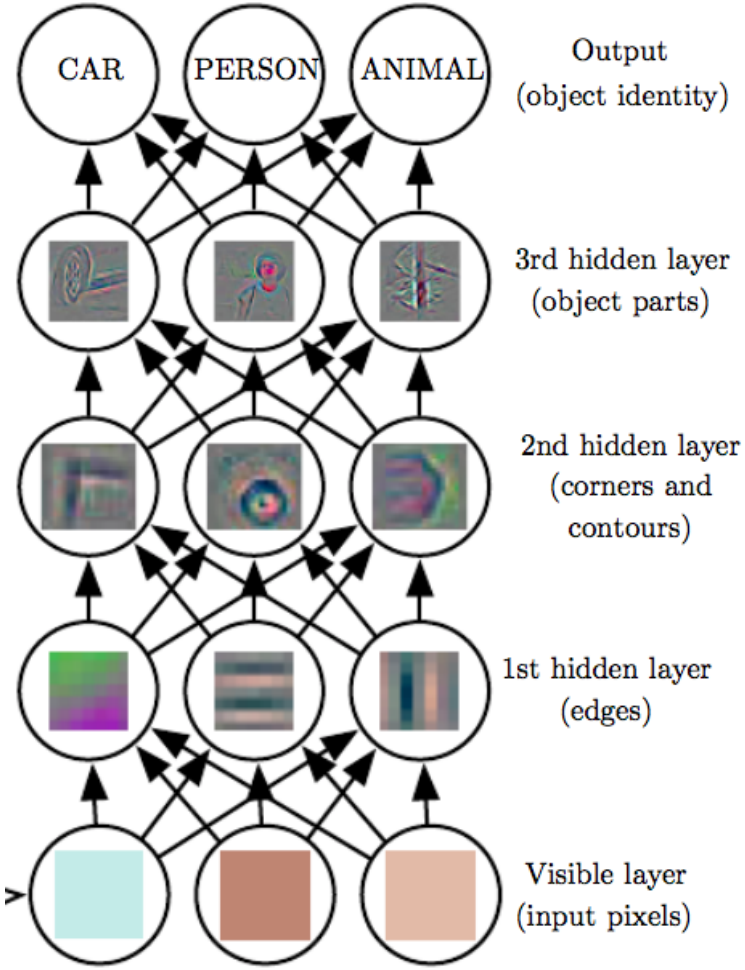


Deep Learning and AI

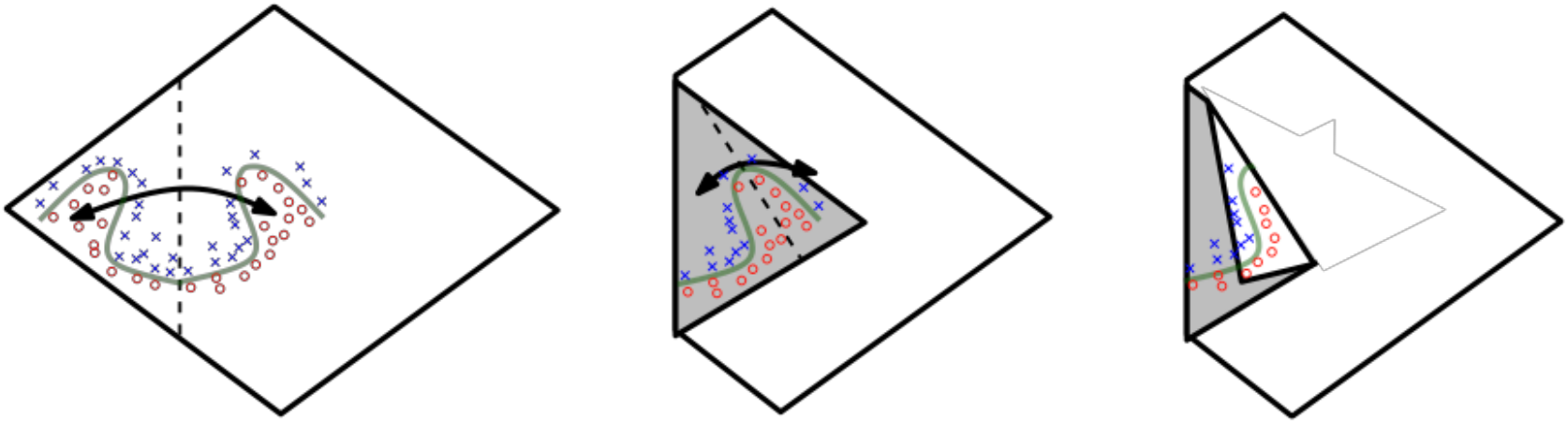


Deep Learning and Representation





Deep Learning and Representation

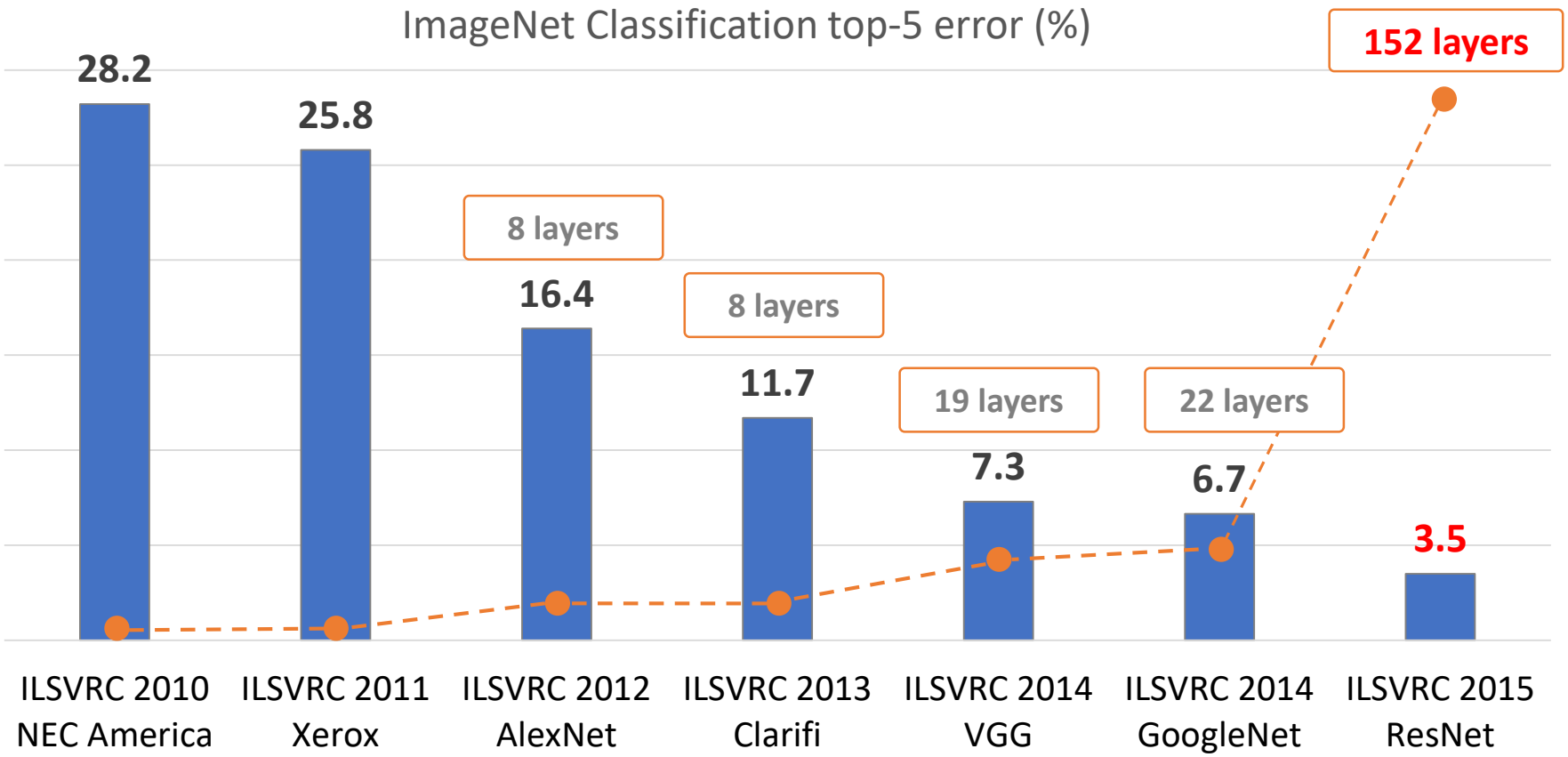


Exponential Representation Advantage of Depth

Why Deep Learning ?

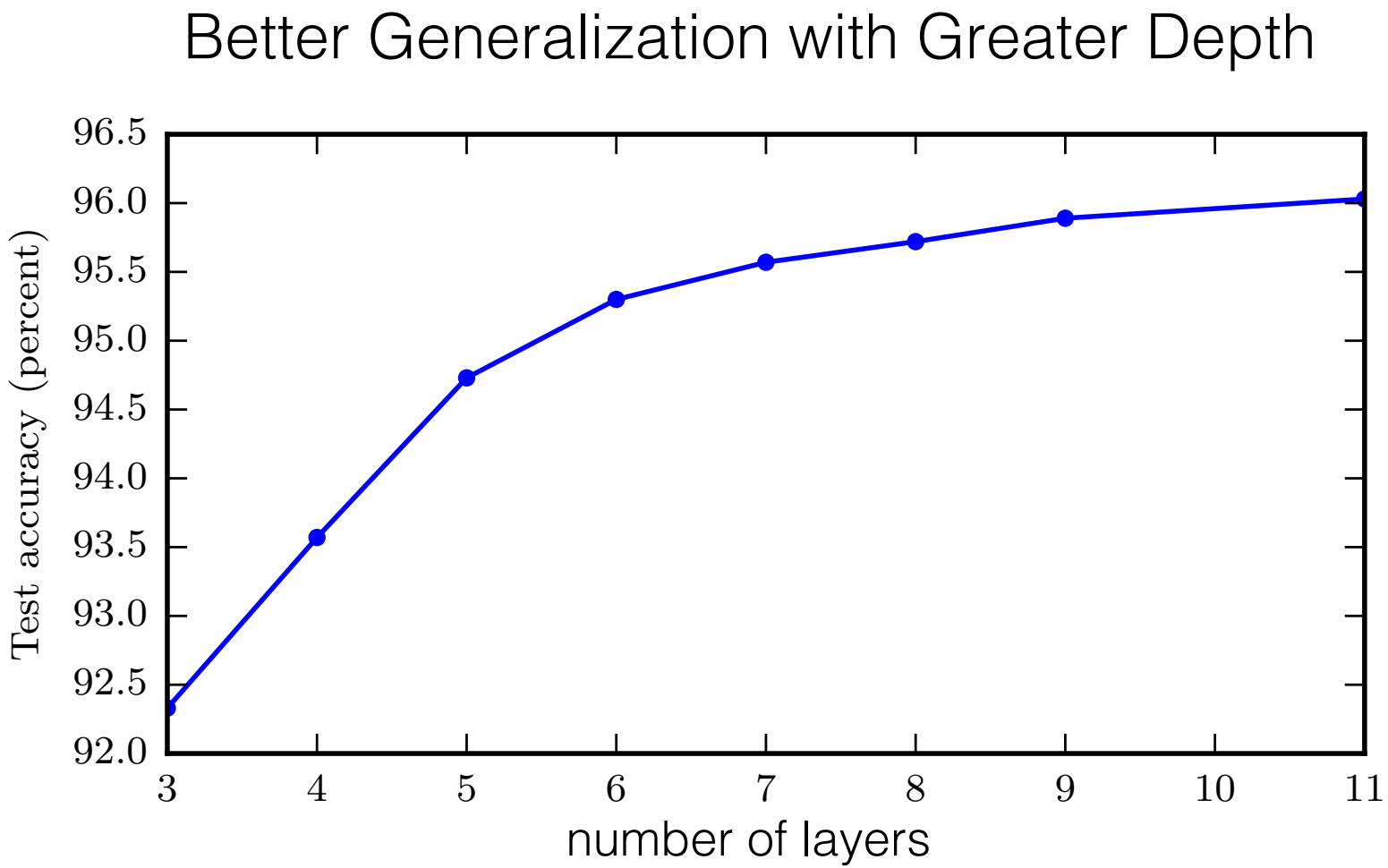
ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

Why Deep Learning ?

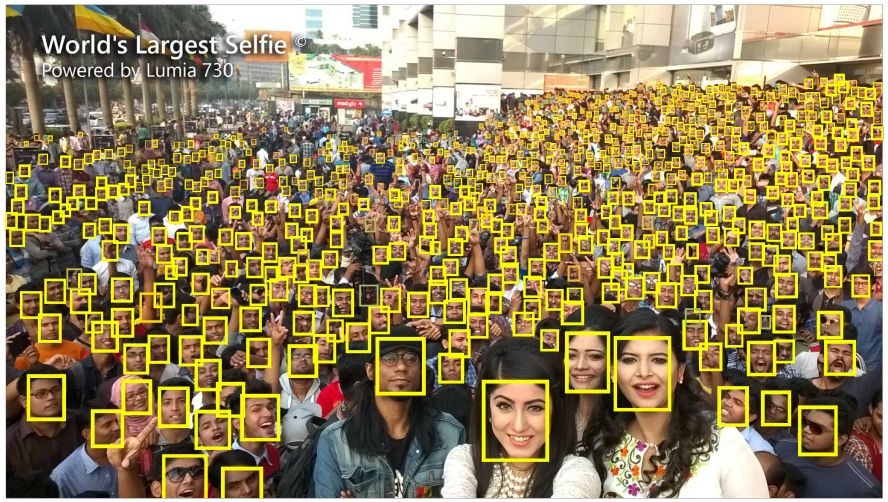


Human error: 5%

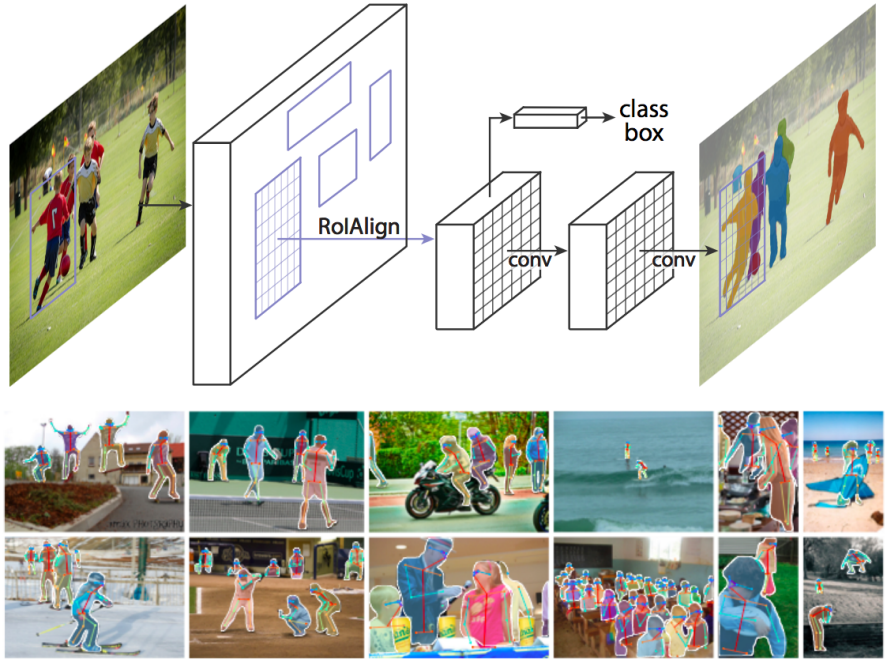
Why Deep Learning ?



Why Deep Learning ?



Hu et al.: Finding tiny faces. 2016



He et al.: Mask R-CNN. 2017

Why Deep Learning ?

Task	Test set	Metric	Best non-neural	Best neural	Source
Machine Translation	Enu-deu newstest16	BLEU	31.4	34.8	http://matrix.statmt.org
	Deu-enu newstest16	BLEU	35.9	39.9	http://matrix.statmt.org
Sentiment Analysis	Stanford sentiment bank	5-class Accuracy	71.0	80.7	Socher+ 13
Question Answering	WebQuestions test set	F1	39.9	52.5	Yih+ 15
Entity Linking	Bing Query Entity Linking set	AUC	72.3	78.2	Gao+ 14b
Image Captioning	COCO 2015 challenge	Turing test pass%	25.5	32.2	Fang+ 15
Sentence compression	Google 10K dataset	F1	0.75	0.82	Fillipova+ 15
Response Generation	Sordoni dataset	BLEU-4	3.98	5.82	Li+ 16a

Task	Test set	Metric	Best non-neural	Best neural	Source
POS tagging	PTB section 23	F1	97.17	97.78	Andor+ 16
Syntactic Parsing	PTB section 23	F1	90.1	93.3	Dyer+ 16
Dependency parsing	PTB section 23	F1	93.22	94.61	Andor+ 16
CCG parsing	CCGBank test	F1	85.2	88.7	Lee+ 16
Inference (NLI)	Stanford NLI corpus	Accuracy	78.2	88.3	Chen+ 16

Deep Learning

- Neural Networks with many hidden layers (deep networks)
  - Insufficient depth can hurt
  - The brain has a deep architecture
  - Cognitive processes seem deep
- New types of units, Dropout, Batch Normalization, Deep Convolutional Networks, Probabilistic Models
- Use of unsupervised learning (autoencoders) for incremental training (one layer at a time) and pre-training in general
- Computational tools
  - Computational graph
- Brute force training using GPUs and accelerating cards

Deep Learning

- Neural Networks with many hidden layers (deep networks)
  - Insufficient depth can hurt
  - The brain has a deep architecture
  - Cognitive processes seem deep

- New types of units, Dropout, Batch Normalization, Deep Convolutional Networks, Probabilistic Models

- Use of unsupervised learning (autoencoders) for incremental training (one layer at a time) and pre-training in general

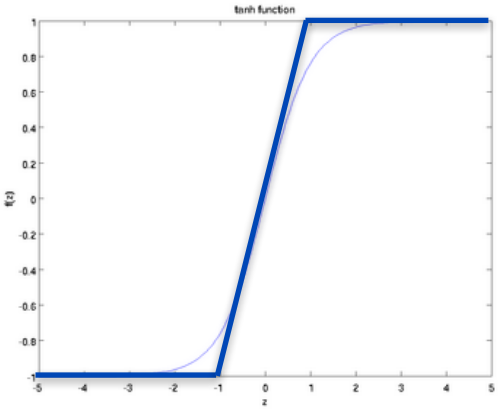
- Computational tools
  - Computational graph

- Brute force training using GPUs and accelerating cards

New Types of Units for Better Gradient Flow

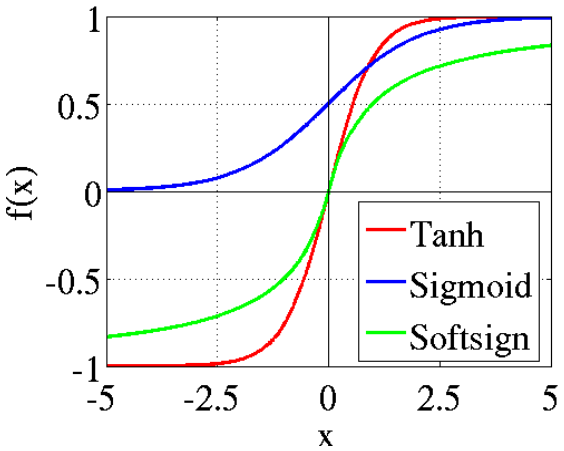
hard tanh

$$\text{HardTanh}(x) = \begin{cases} -1 & \text{if } x < -1 \\ x & \text{if } -1 <= x <= 1 \\ 1 & \text{if } x > 1 \end{cases}$$



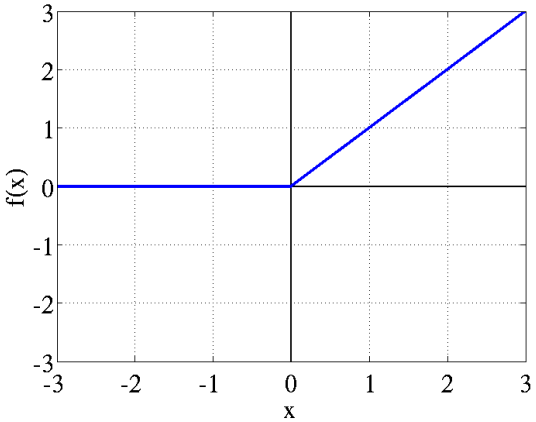
soft sign

$$\text{softsign}(z) = \frac{a}{1+|a|}$$



rectified linear (ReLu)

$$\text{rect}(z) = \max(z, 0)$$

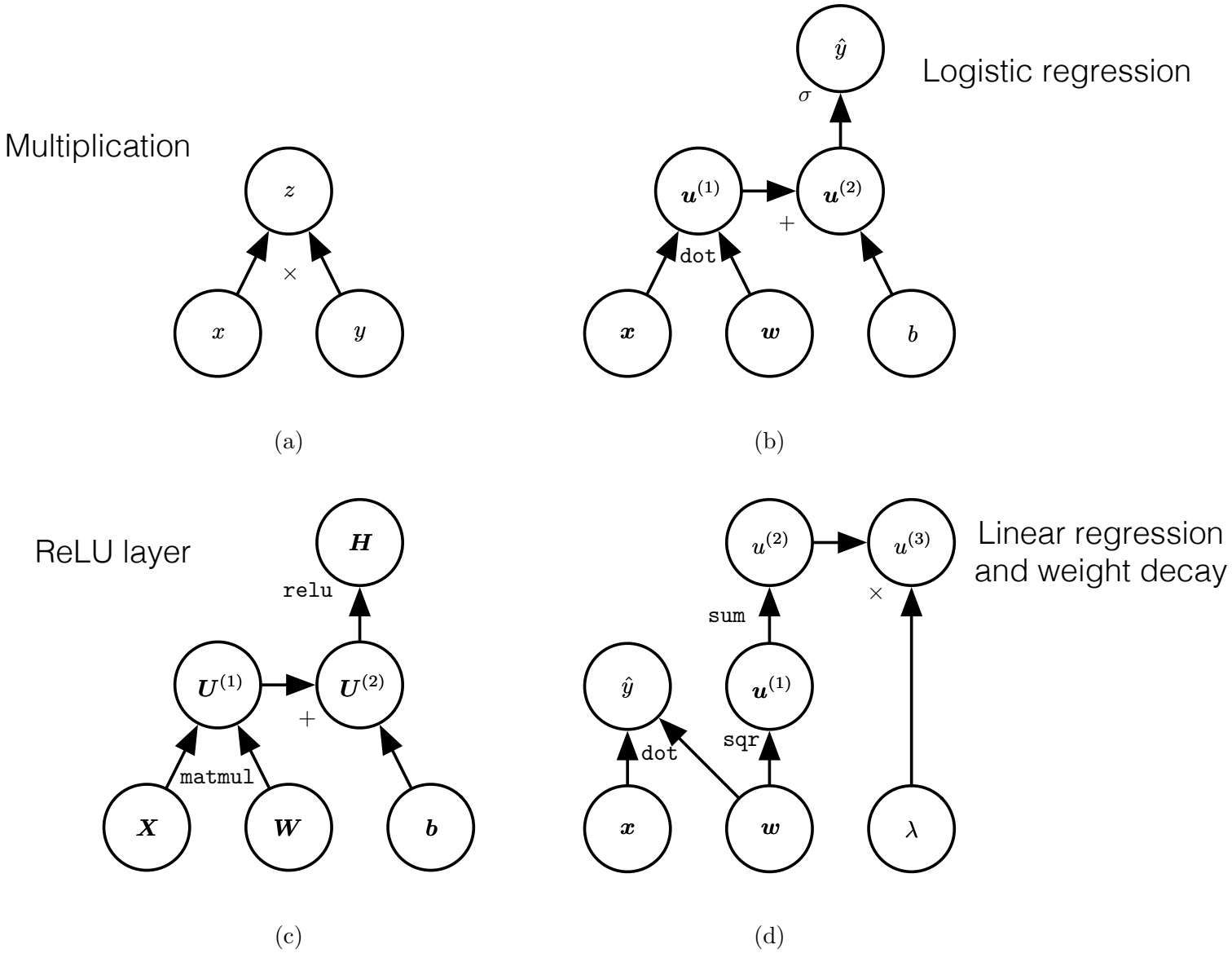


other types of units for Convolution Neural Networks, will see later...

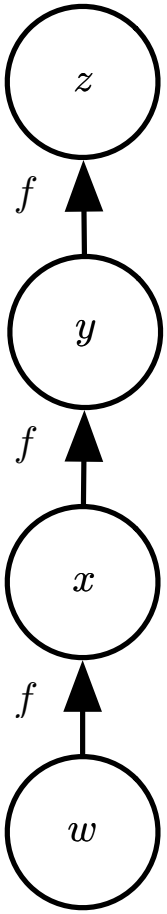
Deep Learning

- Neural Networks with many hidden layers (deep networks)
  - Insufficient depth can hurt
  - The brain has a deep architecture
  - Cognitive processes seem deep
- New types of units, Dropout, Batch Normalization, Deep Convolutional Networks, Probabilistic Models
- Use of unsupervised learning (autoencoders) for incremental training (one layer at a time) and pre-training in general
- Computational tools
  - Computational graph
- Brute force training using GPUs and accelerating cards

Computational Graph



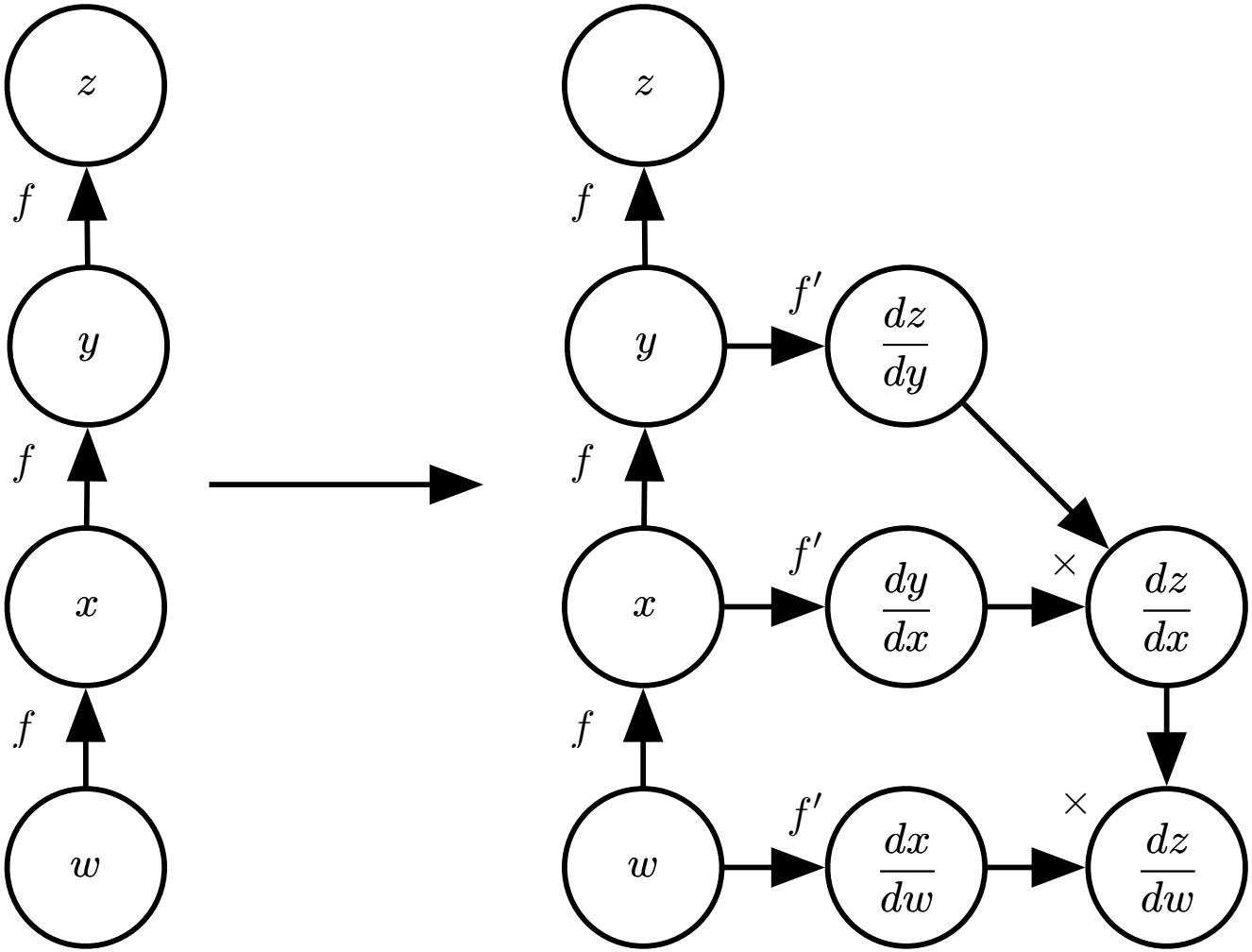
Computational Graph



$$\frac{\partial z}{\partial w}$$
$$= \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}\frac{\partial x}{\partial w}$$
$$= f'(y)f'(x)f'(w)$$
$$= f'(f(f(w)))f'(f(w))f'(w)$$

Back-prop avoids computing this twice

Symbol-to-symbol Differentiation



Loss Function as Computational Graph