Applications

Lecture slides for Chapter 12 of *Deep Learning*

www.deeplearningbook.org

Ian Goodfellow

2016-09-27

Large Scale Deep Learning

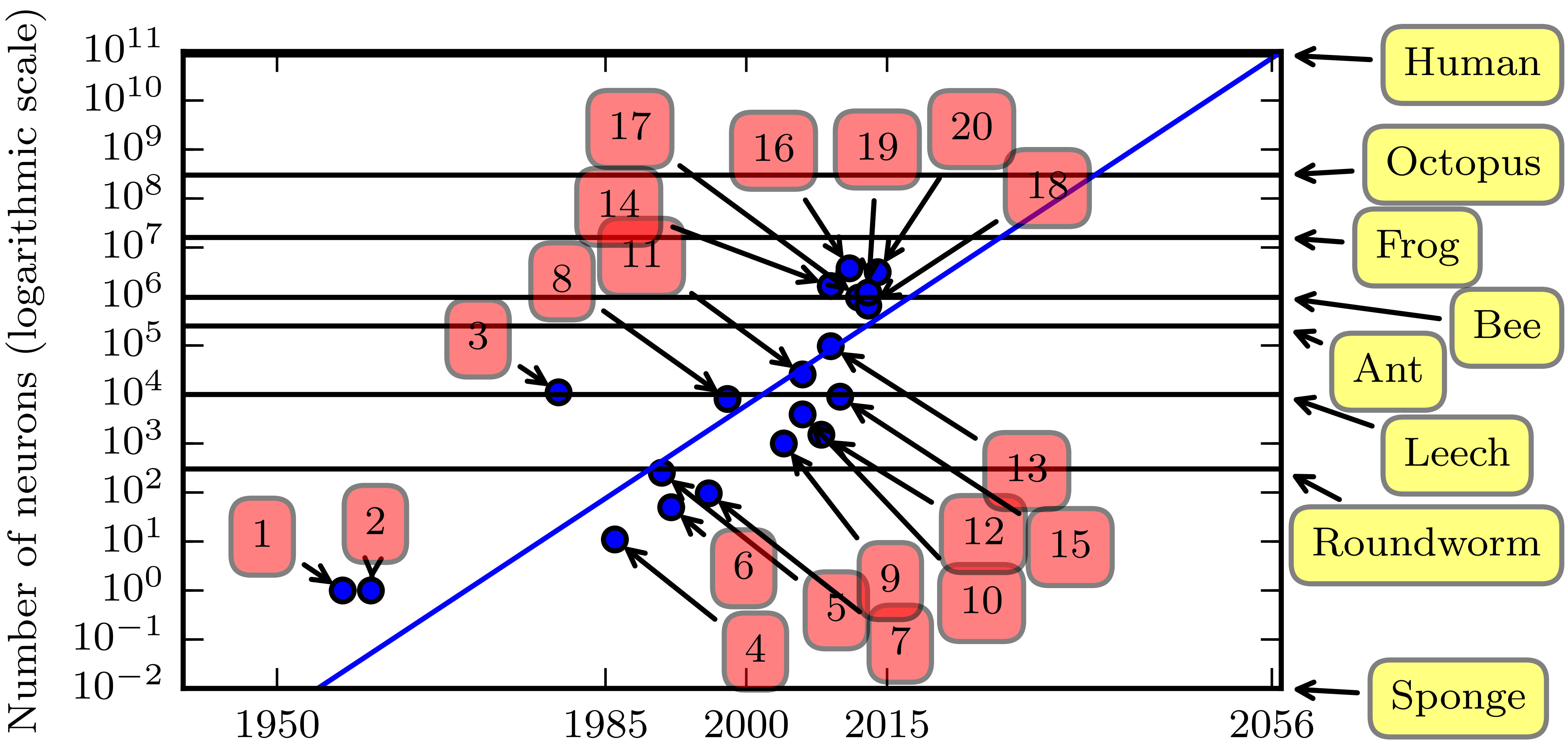


Figure 1.11

(Goodfellow 2016)

Fast Implementations

- CPU
    - Exploit fixed point arithmetic in CPU families where this offers a speedup
    - Cache-friendly implementations
- GPU
    - High memory bandwidth
    - No cache
    - Warps must be synchronized

Distributed Implementations

- Distributed
  - Multi-GPU
  - Multi-machine
    - Model parallelism
    - Data parallelism
      - Trivial at test time
      - Asynchronous SGD at train time

Model Compression

- Large models often have lower test error
  - Very large model trained with dropout
  - Ensemble of many models
- Want small model for low resource use at test time
- Train a small model to mimic the large one
  - Obtains better test error than directly training a small model

Dynamic Structure

- Cascade
- Hard mixture of experts

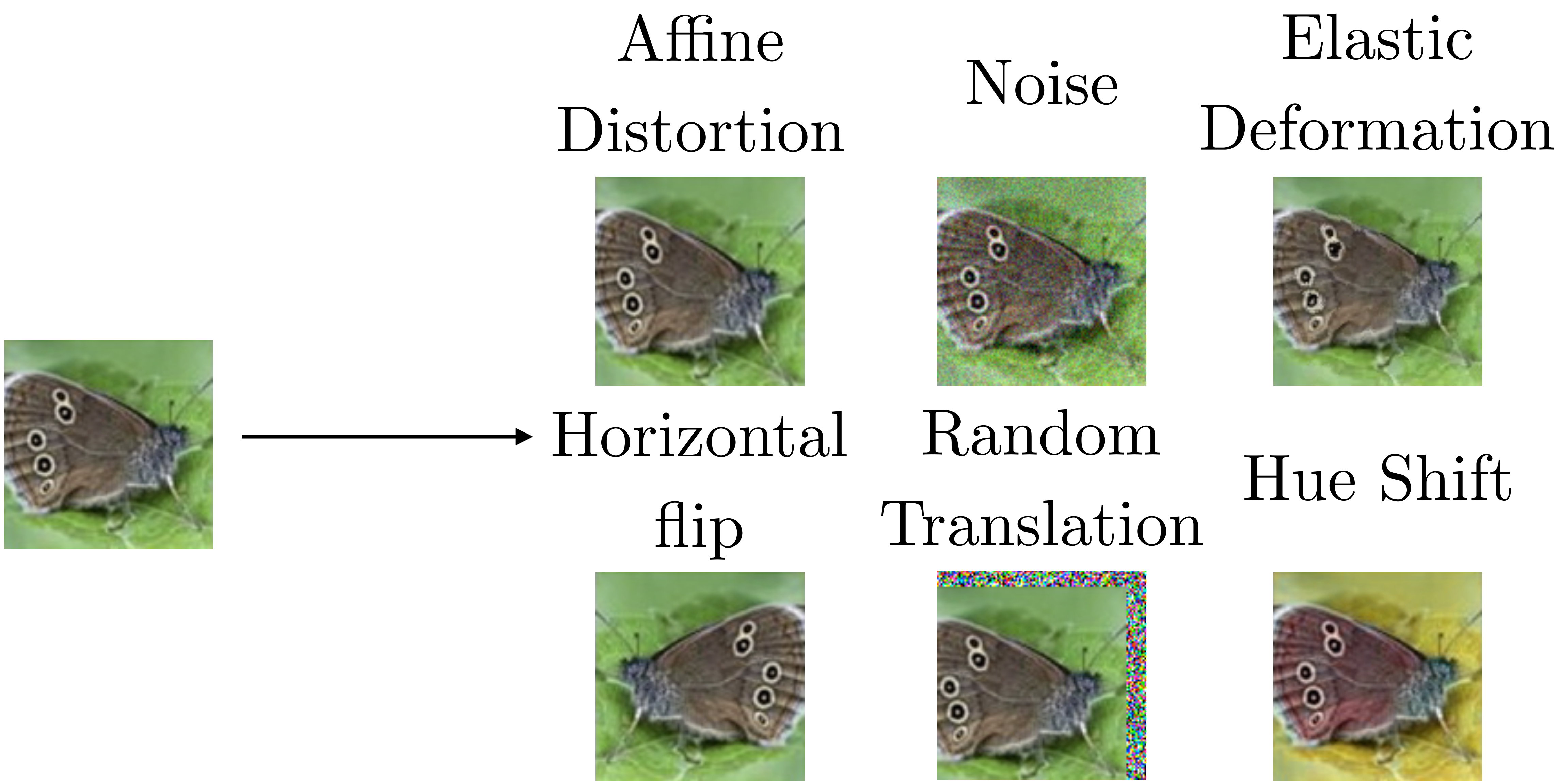
Specialized Hardware Implementations

- ASICs
- Algorithms need to work with lower precision

Computer Vision

- See the convolutional networks chapter
- Specialized preprocessing is no longer very important

Dataset Augmentation for Computer Vision



(Goodfellow 2016)

Speech

- Recognition:
  - LSTMs have replaced HMM-GMMs
- Synthesis:
  - WaveNet

Natural Language Processing

- An important predecessor to deep NLP is the family of models based on n-grams:

$$P(x_1, \ldots, x_\tau) = P(x_1, \ldots, x_{n-1}) \prod_{t=n}^{\tau} P(x_t \mid x_{t-n+1}, \ldots, x_{t-1}). \tag{12.5}$$

$$P(\texttt{THE DOG RAN AWAY}) = P_3(\texttt{THE DOG RAN}) P_3(\texttt{DOG RAN AWAY}) / P_2(\texttt{DOG RAN}). \tag{12.7}$$

Improve with:

-Smoothing

-Backoff

-Word categories

(Goodfellow 2016)

Word Embeddings in Neural Language Models

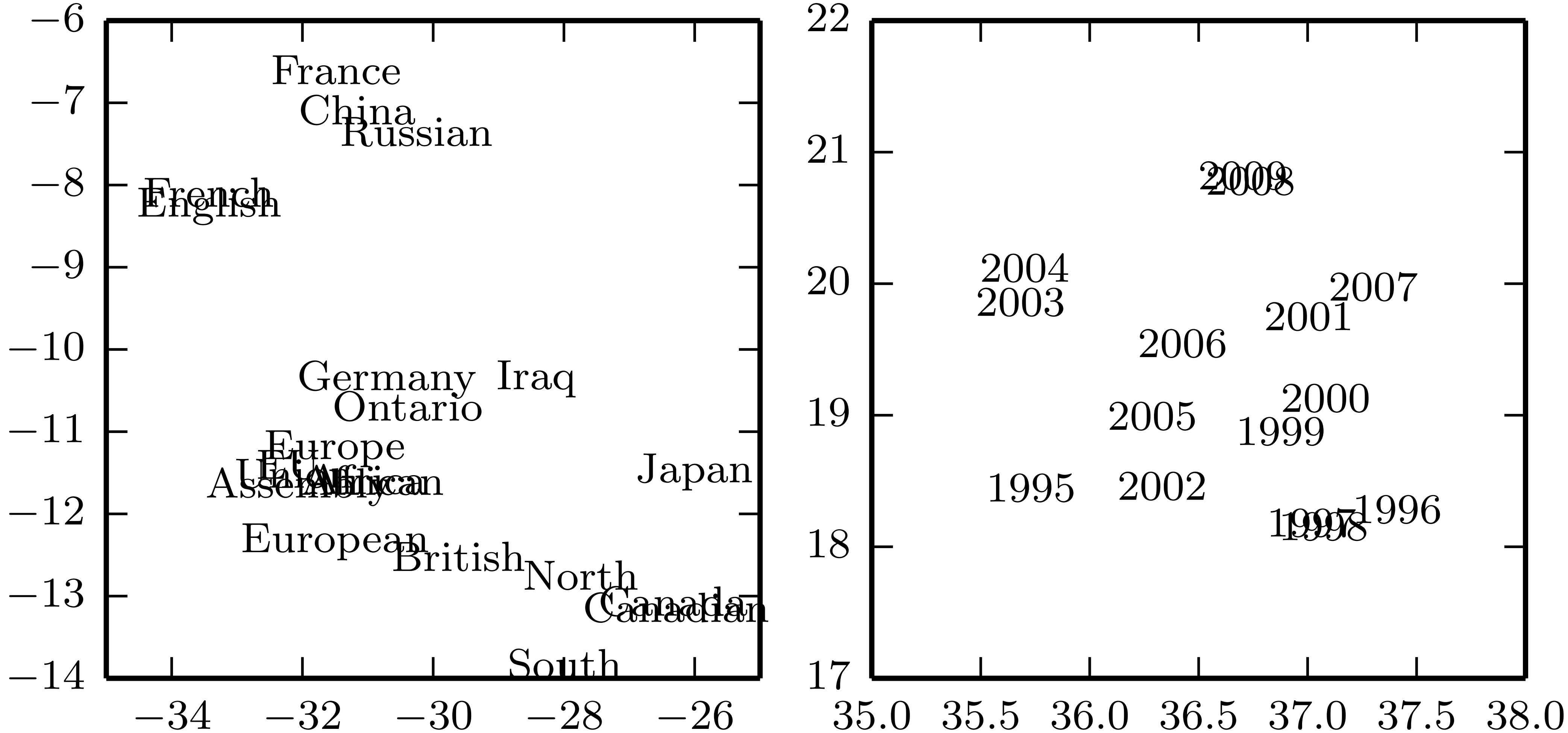


Figure 12.3

(Goodfellow 2016)

High-Dimensional Output Layers for Large Vocabularies

- Short list
- Hierarchical softmax
- Importance sampling
- Noise contrastive estimation

A Hierarchy of Words and Word Categories

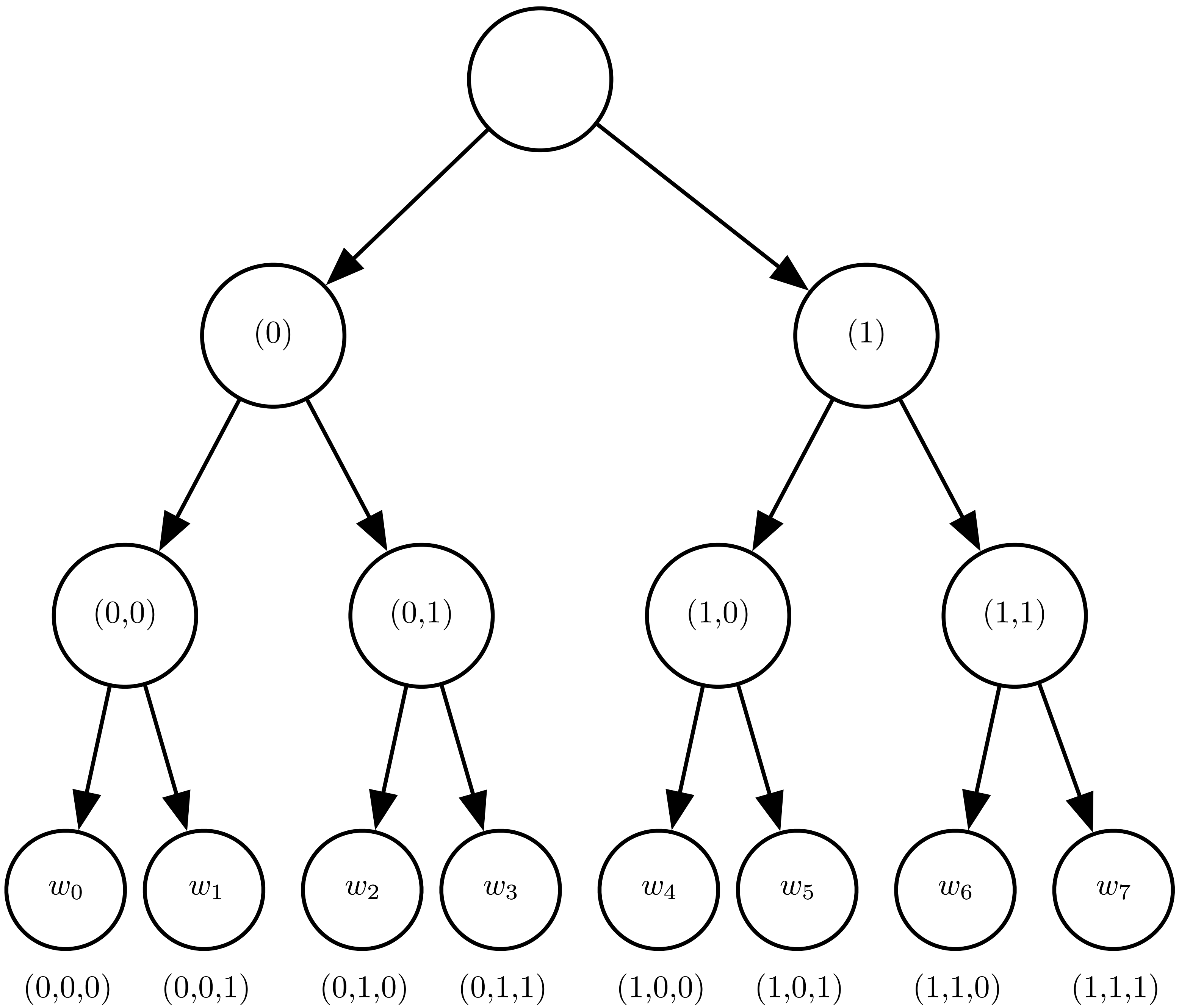


Figure 12.4

(Goodfellow 2016)

Neural Machine Translation

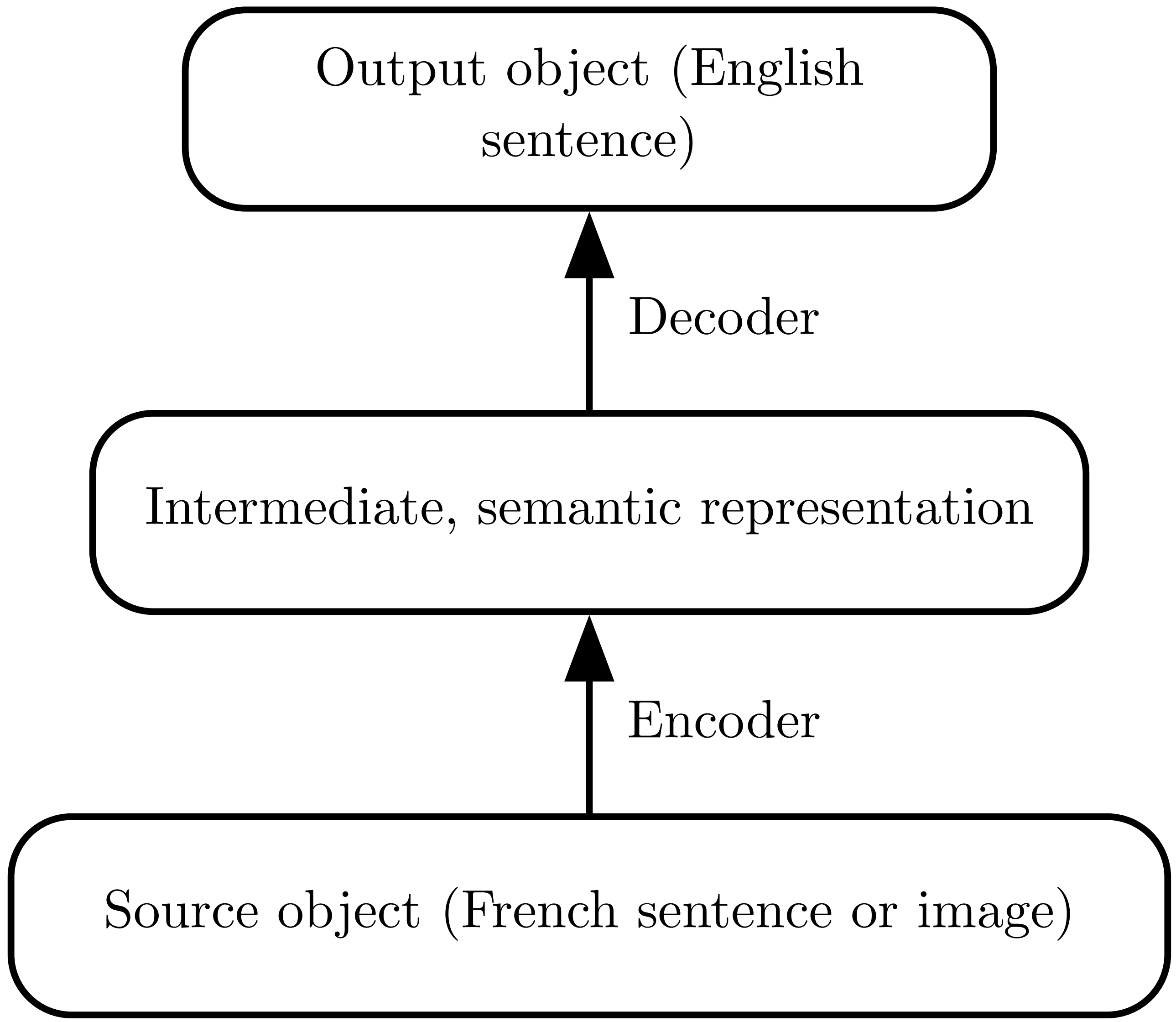


Figure 12.5

(Goodfellow 2016)

Alignment via Attention

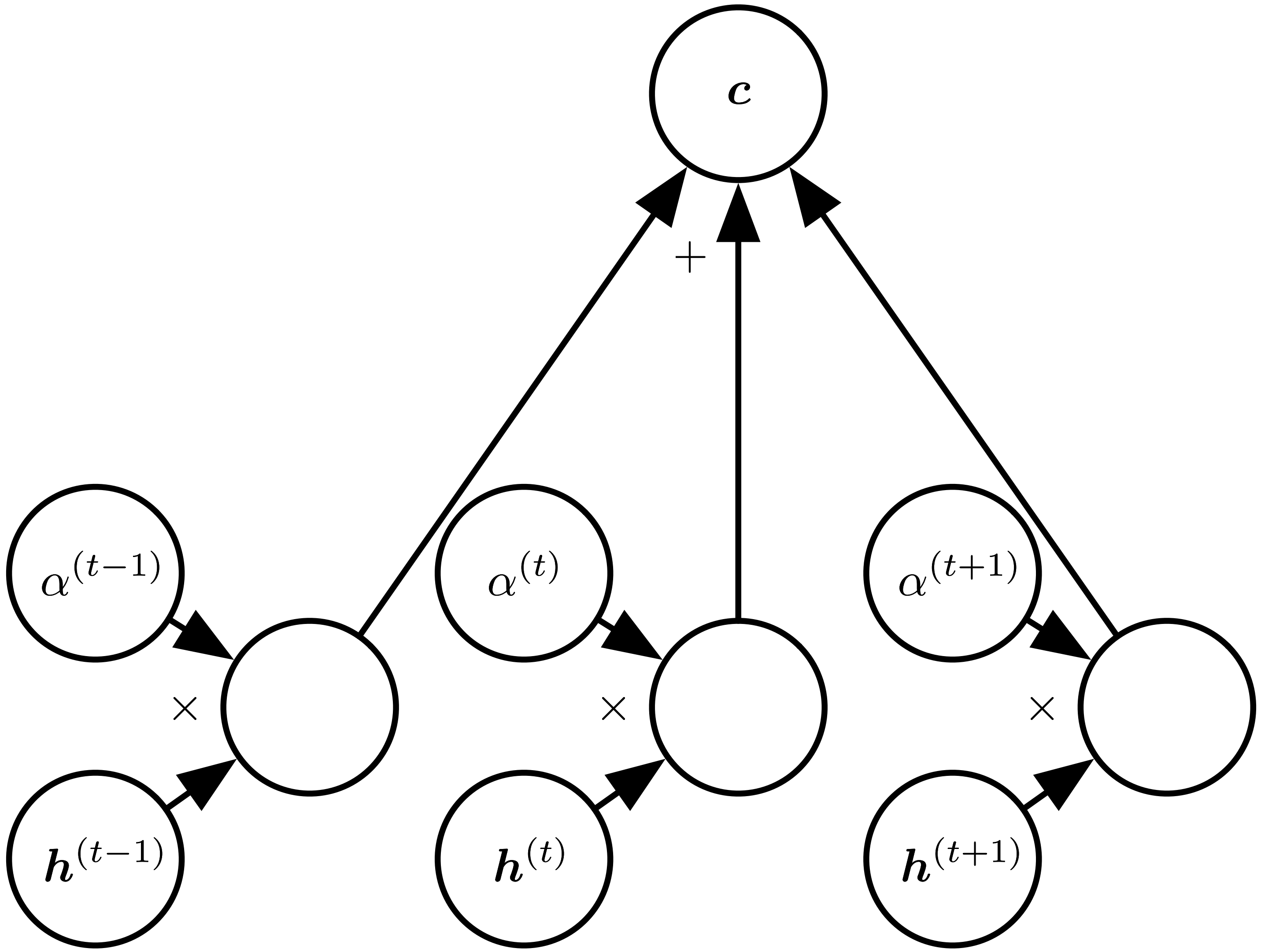


Figure 12.6