Machine Learning – Lecture 21

Wrapping Up

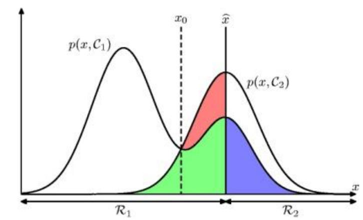
22.01.2019

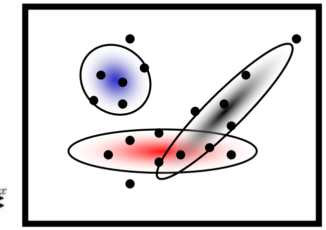
Bastian Leibe
RWTH Aachen
http://www.vision.rwth-aachen.de

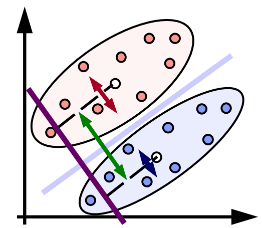
leibe@vision.rwth-aachen.de

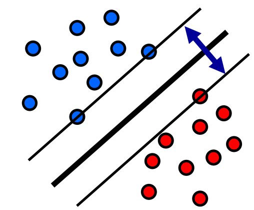
Course Outline

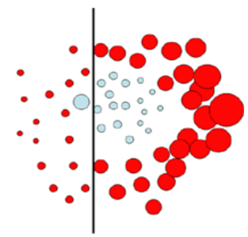
- Fundamentals
  - Bayes Decision Theory
  - Probability Density Estimation
- Classification Approaches
  - Linear Discriminants
  - Support Vector Machines
  - Ensemble Methods & Boosting
  - Random Forests
- Deep Learning
  - Foundations
  - Convolutional Neural Networks
  - Recurrent Neural Networks
  - Current Research Directions

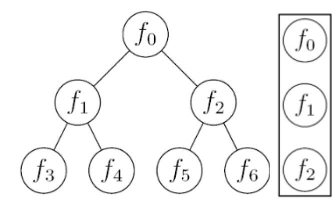


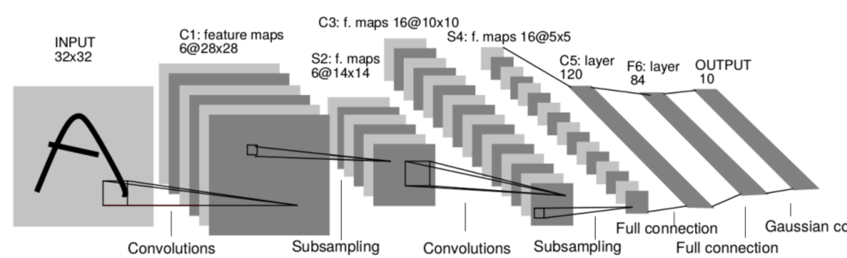




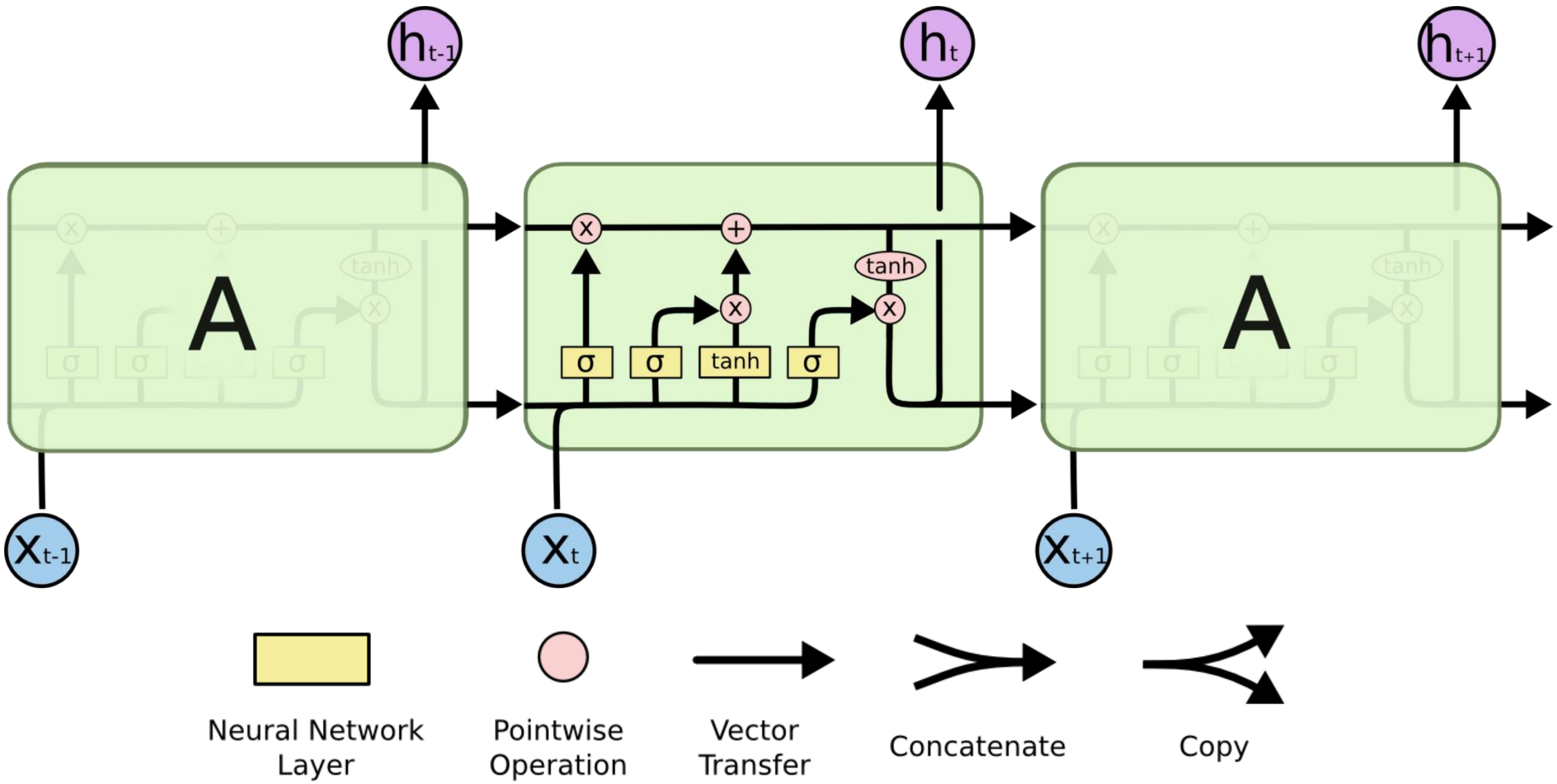








Recap: Long Short-Term Memory



- LSTMs
  - Inspired by the design of memory cells
  - Each module has 4 layers, interacting in a special way.

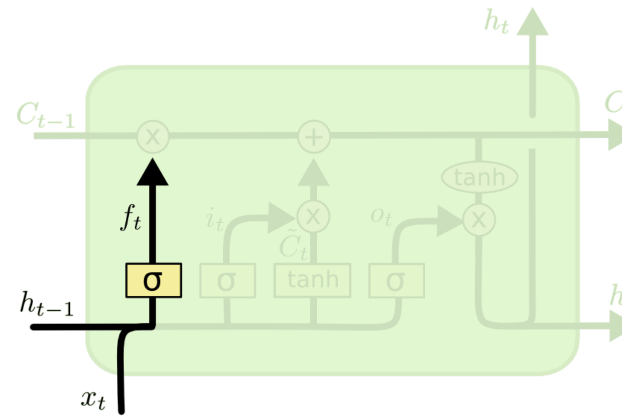
Image source: Christopher Olah, http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Recap: Elements of LSTMs

- Forget gate layer
  - Look at $\mathbf{h}_{t\text{-}1}$ and $\mathbf{x}_t$ and output a number between 0 and 1 for each dimension in the cell state $\mathbf{C}_{t\text{-}1}$.

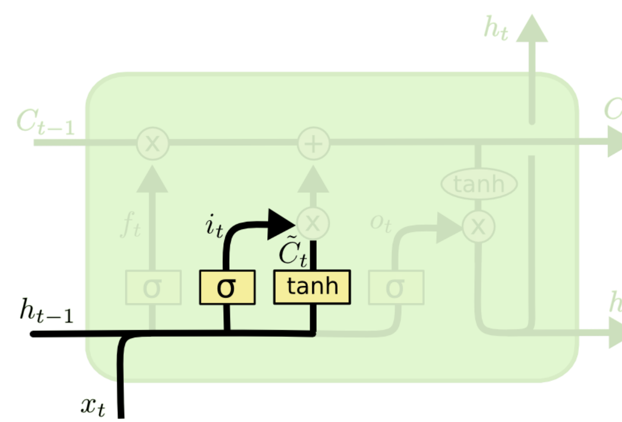
    0: completely delete this,

    1: completely keep this.



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$

- Update gate layer
  - Decide what information to store in the cell state.
  - Sigmoid network (input gate layer) decides which values are updated.
  - tanh layer creates a vector of new candidate values   that could be added to the state.



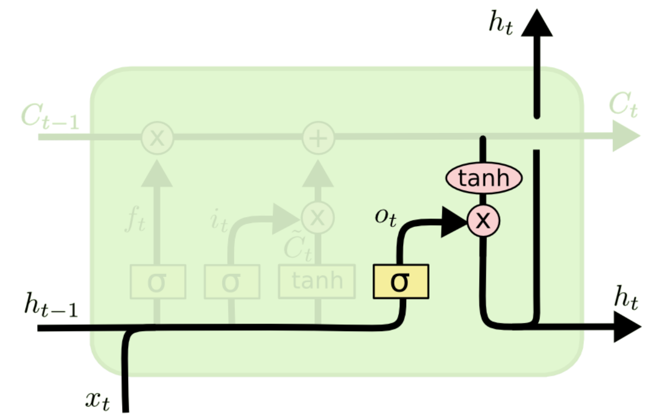
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$

Source: Christopher Olah, http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Recap: Elements of LSTMs

- Output gate layer
  - Output is a filtered version of our gate state.
  - First, apply sigmoid layer to decide what parts of the cell state to output.
  - Then, pass the cell state through a tanh (to push the values to be between -1 and 1) and multiply it with the output of the sigmoid gate.

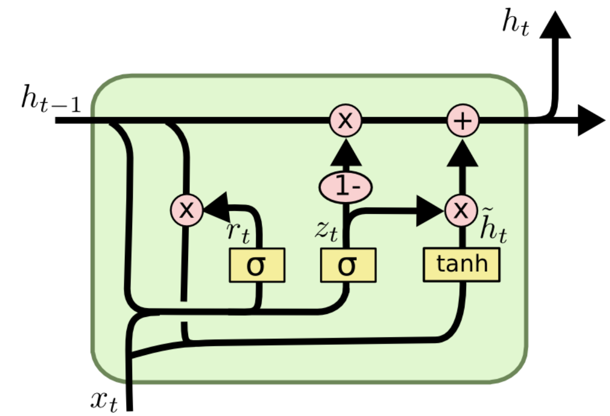


$$o_t = \sigma\left(W_o\ [h_{t-1}, x_t]\ +\ b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

Source: Christopher **Olah, http://colah.github.io/posts/2015-08-Understanding-LSTMs/**

Recap: Gated Recurrent Units (GRU)

- Simpler model than LSTM
  - Combines the forget and input gates into a single update gate z_t.
  - Similar definition for a reset gate r_t, but with different weights.
  - In both cases, merge the cell state and hidden state.



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

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

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

- Empirical results
  - Both LSTM and GRU can learn much longer-term dependencies than regular RNNs
  - GRU performance similar to LSTM (no clear winner yet), but fewer parameters.

Source: Christopher Olah, http://colah.github.io/posts/2015-08-Understanding-LSTMs/

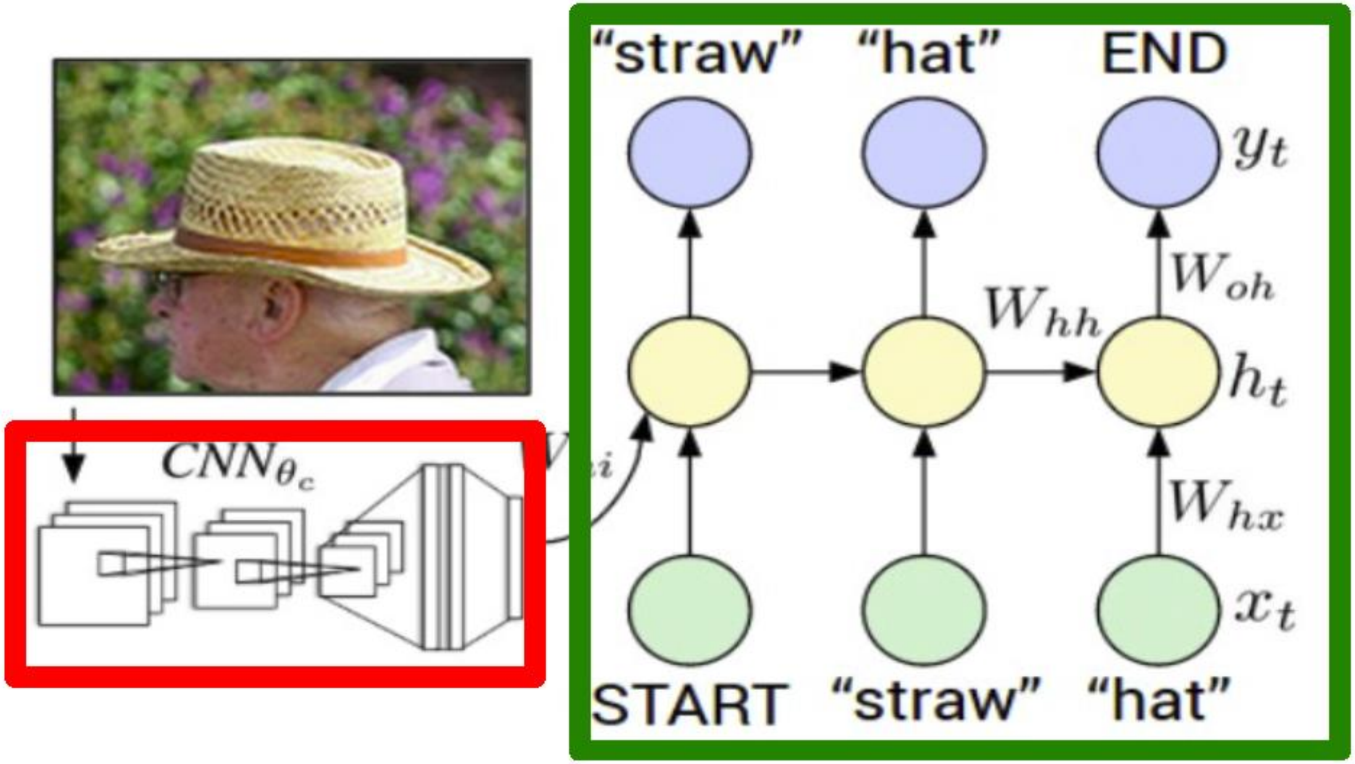
Recap: Language Model Results

```
/*
 * Increment the size file of the new incorrect UI_FILTER group information
 * of the size generatively.
 */
static int indicate_policy(void)
{
  int error;
  if (fd == MARN_EPT) {
    /*
     * The kernel blank will coeld it to userspace.
     */
    if (ss->segment < mem_total)
      unblock_graph_and_set_blocked();
    else
      ret = 1;
    goto bail;
  }
  segaddr = in_SB(in.addr);
  selector = seg / 16;
  setup_works = true;
  for (i = 0; i < blocks; i++) {
    seq = buf[i++];
    bpf = bd->bd.next + i * search;
    if (fd) {
```

- Example:
  Hallucinating C Code
  - Trained on the Linux source code (474MB from github)
  - Using a large 3-layer LSTM

Slide adapted from Andrej Karpathy

Applications: Image Tagging



- Simple combination of CNN and RNN
  - Use CNN to define initial state $\mathbf{h}_0$ of an RNN.
  - Use RNN to produce text description of the image.

Applications: Image Tagging

- Setup
  - Train on corpus of images with textual descriptions
  - E.g. Microsoft CoCo
    - 120k images
    - 5 sentences each

a man riding a bike on a dirt path through a forest.
bicyclist raises his fist as he rides on desert dirt trail.
this dirt bike rider is smiling and raising his fist in triumph.
a man riding a bicycle while pumping his fist in the air.
a mountain biker pumps his fist in celebration.

Results: Image Tagging

a group of people standing around a room with remotes
`logprob: -9.17`

a young boy is holding a baseball bat
`logprob: -7.61`

a cow is standing in the middle of a street
`logprob: -8.84`

Spectacular results!

Results: Image Tagging

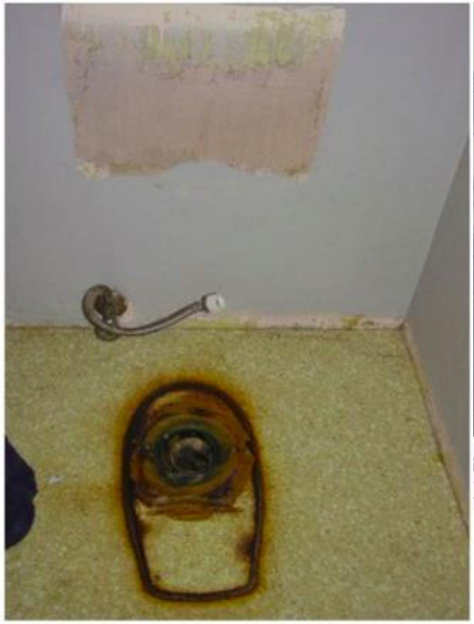
a baby laying on a bed with a stuffed bear
logprob: -8.66

a young boy is holding a baseball bat
logprob: -7.65

a cat is sitting on a couch with a remote control
logprob: -12.45

- Wrong, but one can still see why those results were selected...

Results: Image Tagging



a toilet with a seat up in a bathroom
logprob: -13.44

a woman holding a teddy bear in front of a mirror
logprob: -9.65

a horse is standing in the middle of a road
logprob: -10.34

- Not sure what happened here...

Fun Application: Image to Story

Later on the eighth day , Billy was a friend of a man who lived on his own . He did n't know how far away they were , and if he was to survive the fall . His mind raced , trying not to show any signs of weakness . The wind ruffled the snow and ice in the snow . He had no idea how many times he was going to climb into the mountains . He told me to stay on the ground for a while , but if I find out what s going on , we should go on foot . Sam and Si Lei joined us in the army .

- Example: Generating a story from an image
  - Trained on corpus of adventure novels

Source: Ryan Kiros http://www.cs.toronto.edu/~rkiros/adv_L.html

More Results

Having lain on the bed , I did n't know what to say . He turned his attention to the room and saw a large room . The room was furnished with a single bed , a dresser and a large bed with a table in the center of the room . It was a long time ago . The room was designed with the most powerful and efficient ones . As far as I m concerned , it was a long time ago . On the other side of the room was a beautiful picture of a woman who had been abducted by the fireplace and their own personal belongings in order to keep it safe , but it didn t take too long . Feeling helpless , he turned his attention back to me . ``

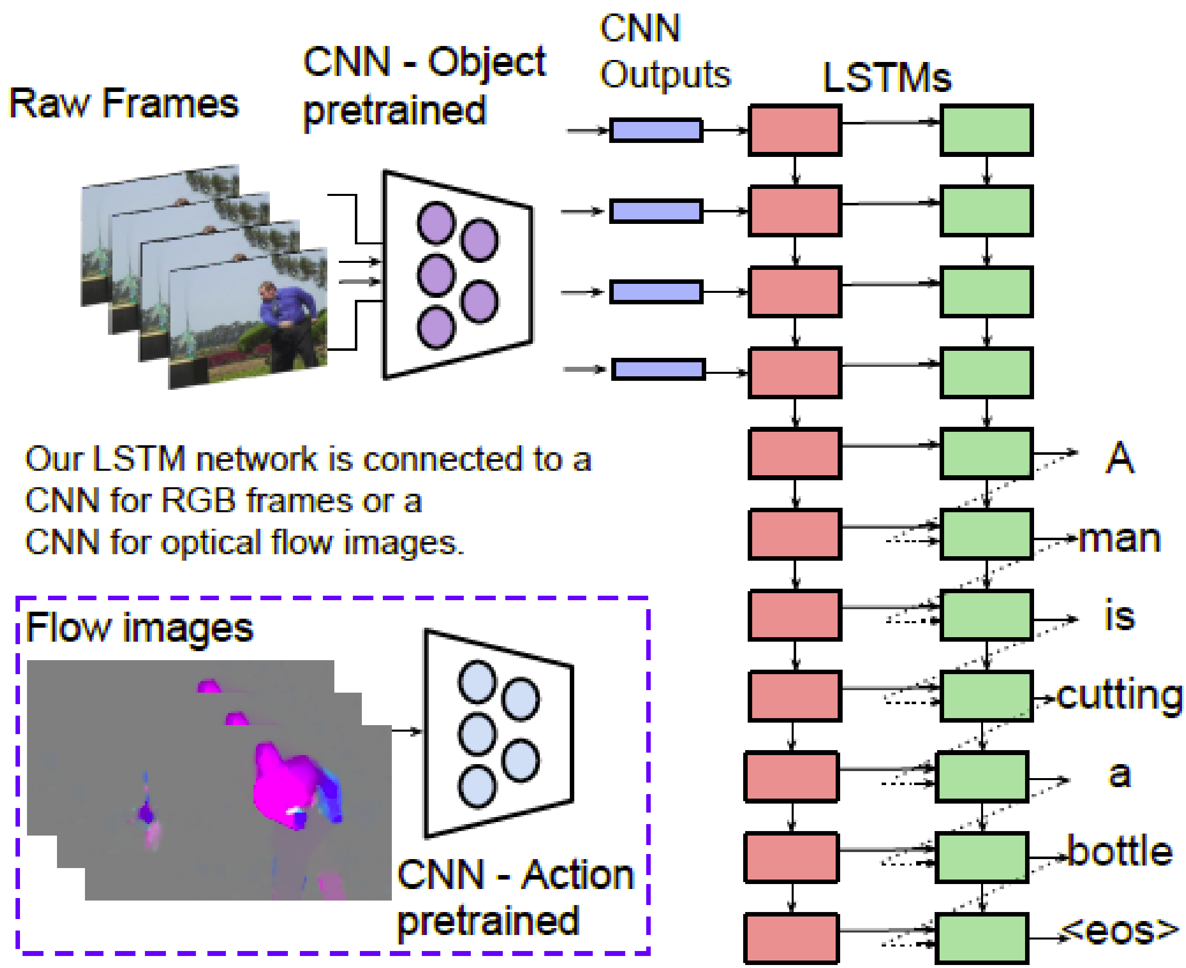
Source: Ryan Kiros http://www.cs.toronto.edu/~rkiros/adv_L.html

More Results

Only Prince Darin knew how to run from the mountains , and once more , he could see the outline of a rider on horseback . The wind ruffled his hair in an attempt to locate the forest . He hadn t been in such a state of mind before , but it was a good thing . All of them seemed to be doing the same thing . They did n't know where they came from . The wind blew up the mountain peaks and disappeared into the sky , leaving trails behind the peaks of the mountains on Mount Fuji .

Source: Ryan Kiros http://www.cs.toronto.edu/~rkiros/adv_L.html

Application: Video to Text Description



Source: Subhashini Venugopalan, ICCV'15

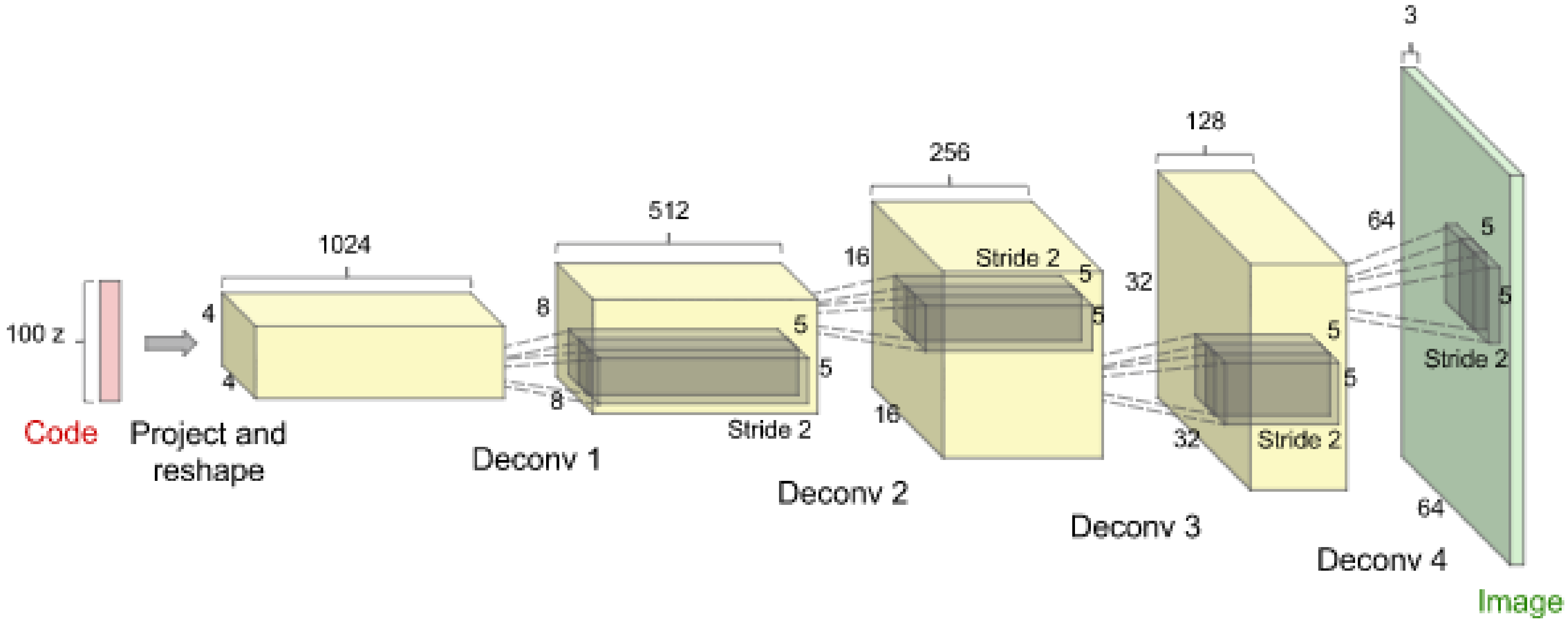
Video-to-Text Results

Source: Subhashini Venugopalan, ICCV’15

Currently Hot Research Directions

- Generative Models
  - Networks for image generation
  - Generative Adversarial Networks (GAN)
- Towards General Models of Computation
  - Memory Networks
  - Neural Turing Machines
- Deep Reinforcement Learning

Generative Networks

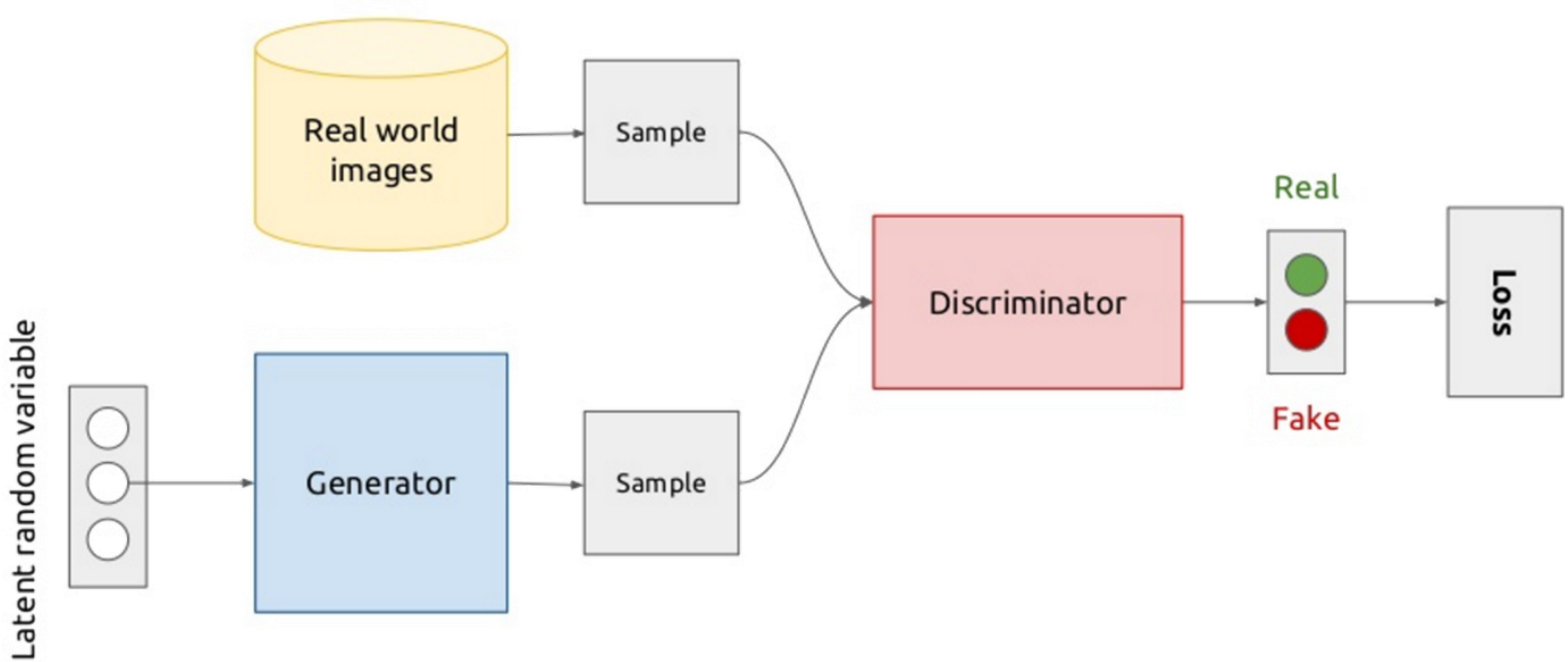


- Using a network to generate images
  - Sampling from noise distribution
  - Sequence of upsampling layers to generate an output image
  - *How can we train such a model to produce the desired output?*

Image from https://blog.openai.com/generative-models/

Generative Adversarial Networks (GAN)

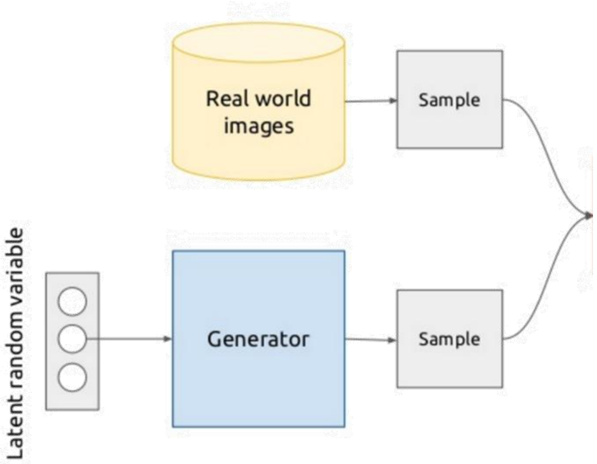
- Conceptual view



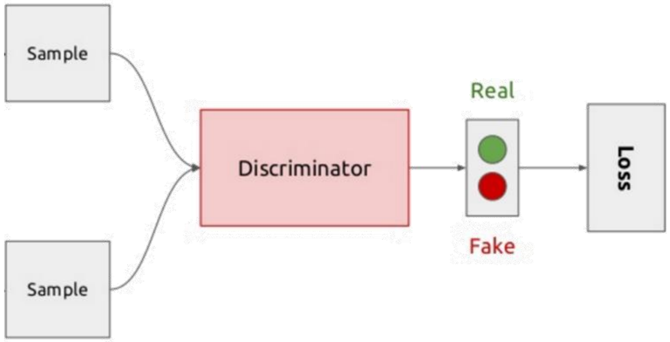
- Main idea
  - Simultaneously train an image generator and a discriminator.
  - Interpreted as a two-player game
  - Very tricky to train…

Two-Player Game

- Generator
  - Tries to draw samples from $p(\boldsymbol{x})$.
  - Analogy: *counterfeiter*



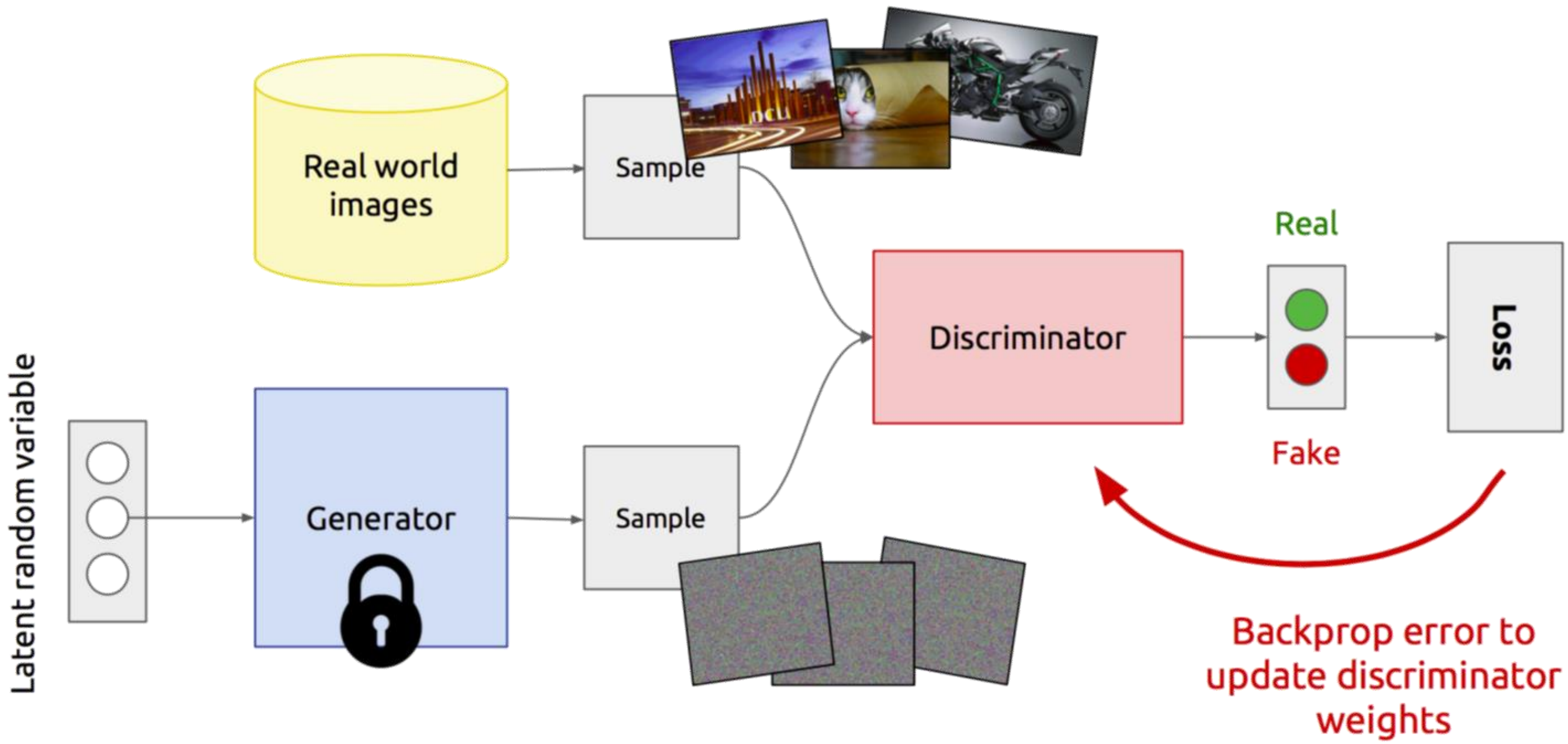
- Discriminator
  - Tries to determine whether the sample came from the generator or the data distribution.
  - Analogy: *police investigator*



- Both generator and discriminator are deep networks
  - We can train them with backprop.

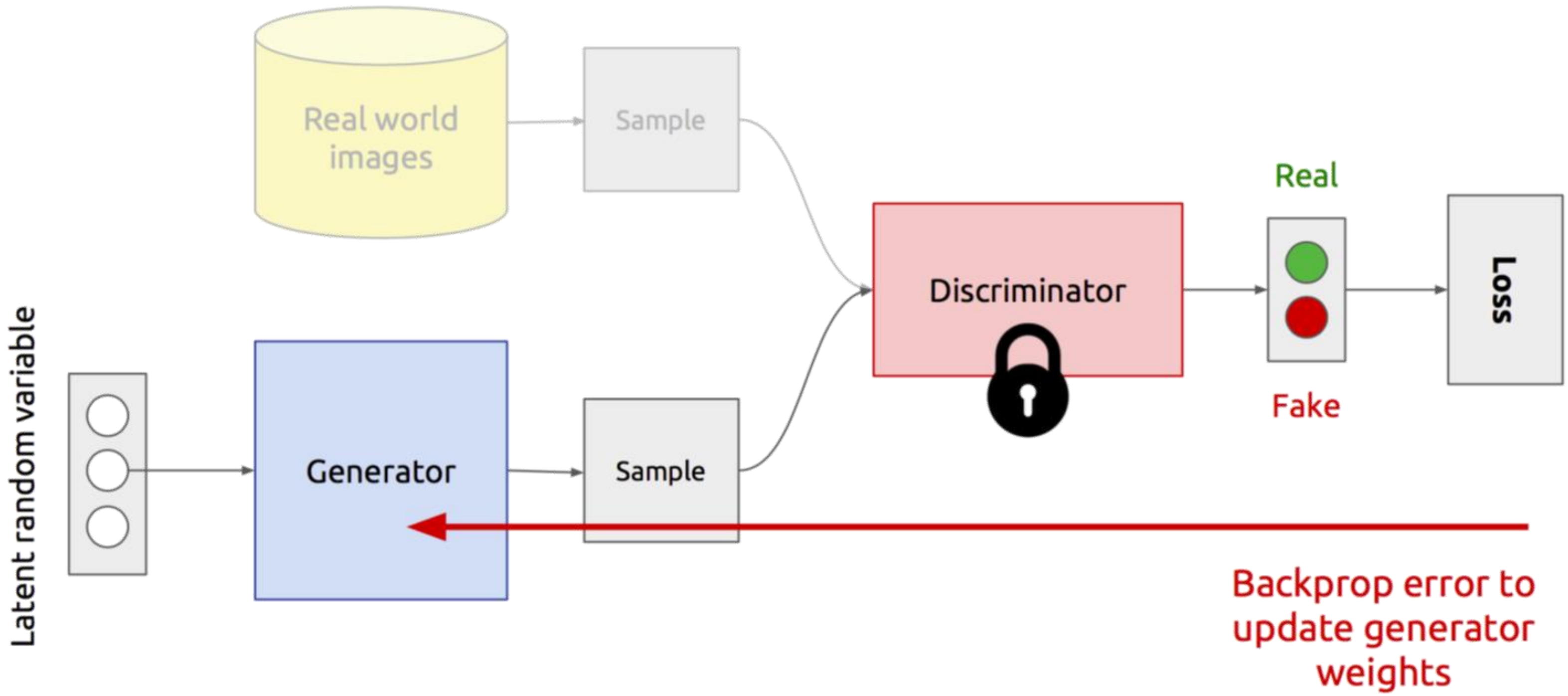
www.bundesbank.de

Training the Discriminator



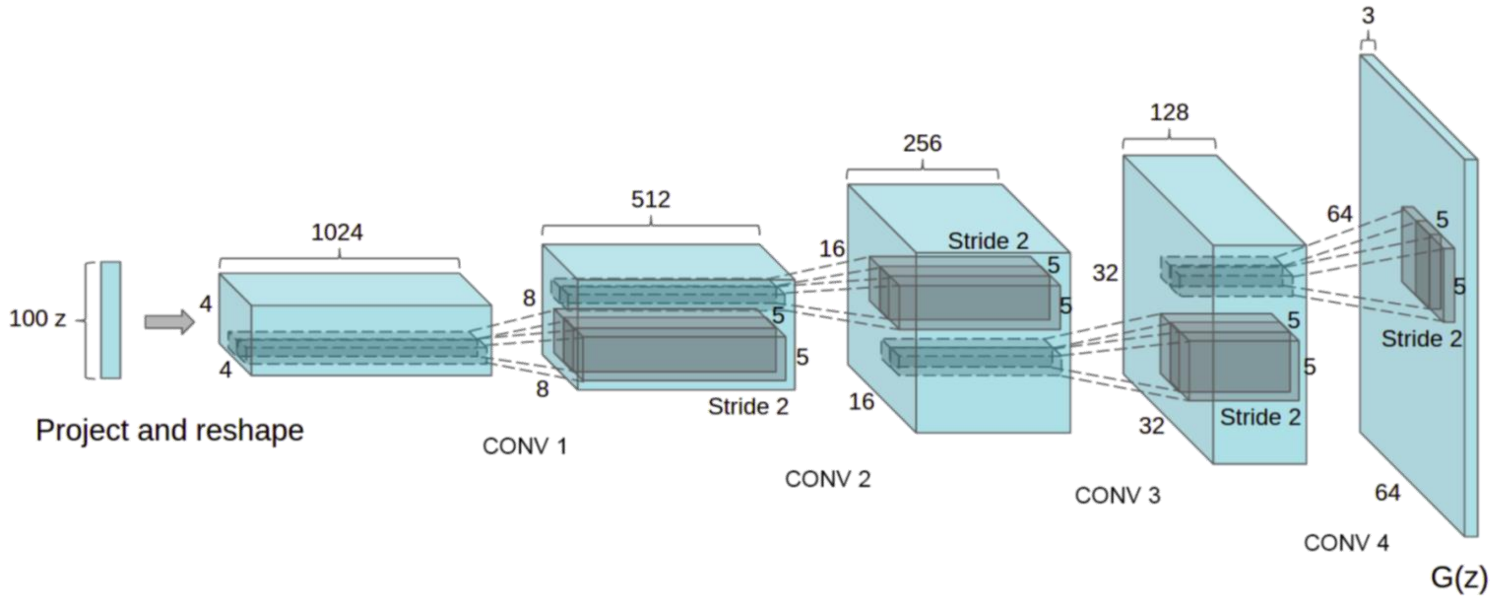
- Procedure
  - Fix generator weights
  - Train discriminator to distinguish between real and generated images

Training the Generator



- Procedure
  - Fix discriminator weights
  - Sample from generator
  - Backprop through discriminator to update generator weights

Example: Deep Convolutional GAN (DCGAN)



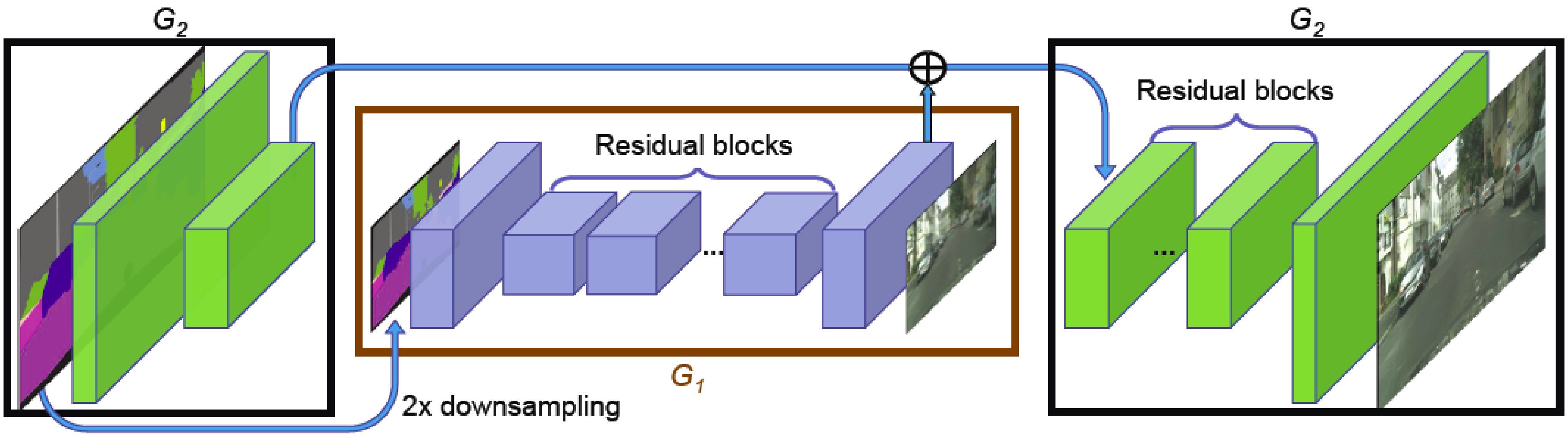
- Generator for images
  - Remove fully-connected layers
  - Upsampling with fractional strided convolutions
  - Batch normalization after each layer (important!)
  - Use ReLu in generator for hidden layers, tanh for output layer
  - Use Leaky ReLu in the discriminator for all layers

Example Application: Image Generation

- Generating bedroom images
  - Each sample is generated from a sampled random number

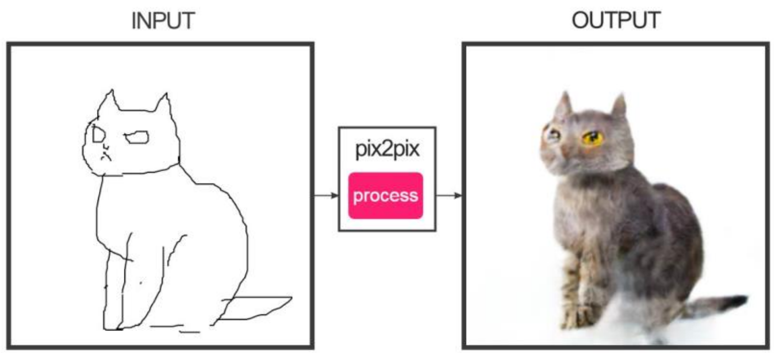
Example Application: Super-Resolution (SRGAN)

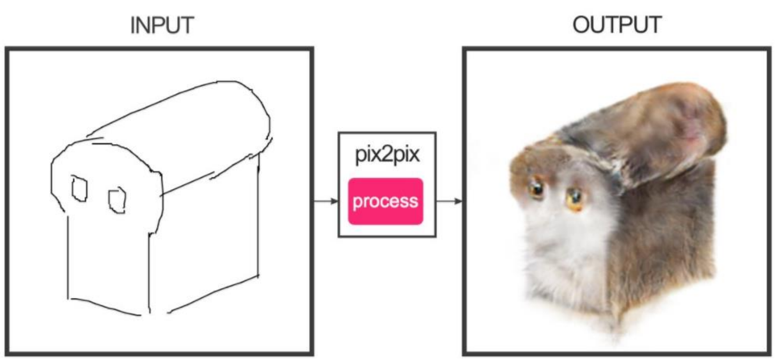
Extension: Conditional GANs



- Idea
  - Condition the latent space representation on an input image
  - Used to create the pix2pix network

Artist Project: edges2cats [Christopher Hesse]





Ivy Tasi @ivymyt

Vitaly Vidmirov @vvid

@ka92

Slide credit: Aljosa Efros

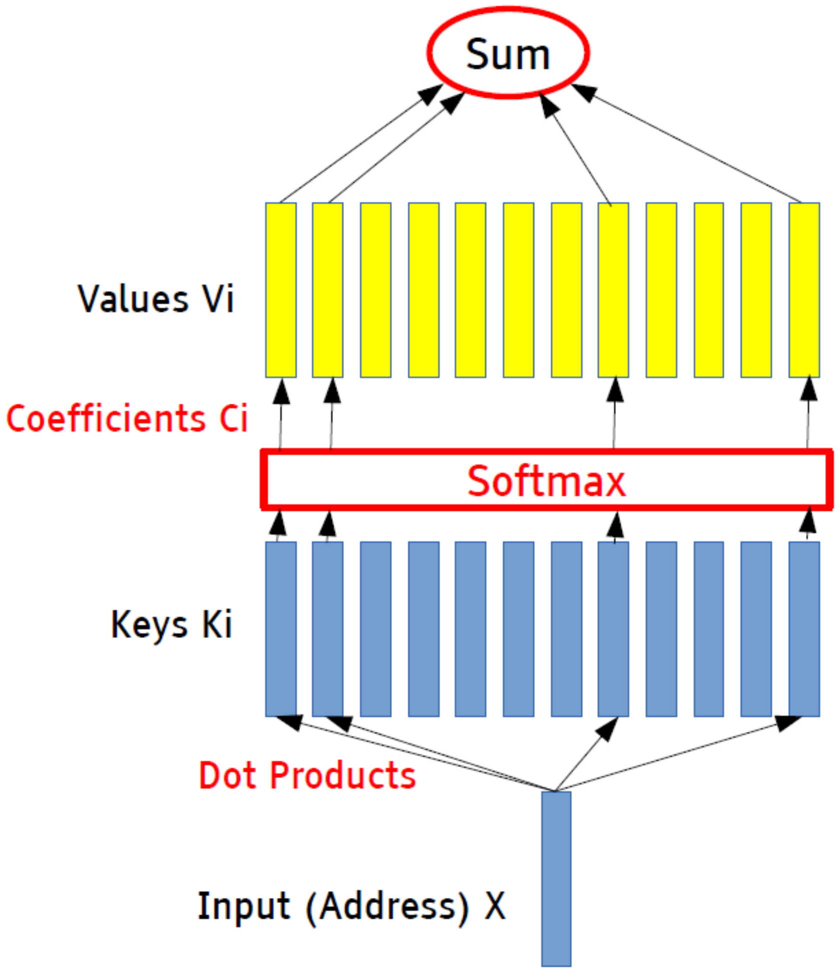
Currently Hot Research Directions

- Generative Models
  - Networks for image generation
  - Generative Adversarial Networks (GAN)
- Towards General Models of Computation
  - Memory Networks
  - Neural Turing Machines
- Deep Reinforcement Learning

Memory Networks

- Soft, differentiable memory
  - Stores <key, value> pairs
  - Input is matched to the stored keys
  - Output is the average over all values that correspond to the matched keys

- Key Idea
  - Make all steps differentiable.
  - ⇒ Then all parameters (including access keys, stored values, etc.) can be learned with end-to-end supervised learning.

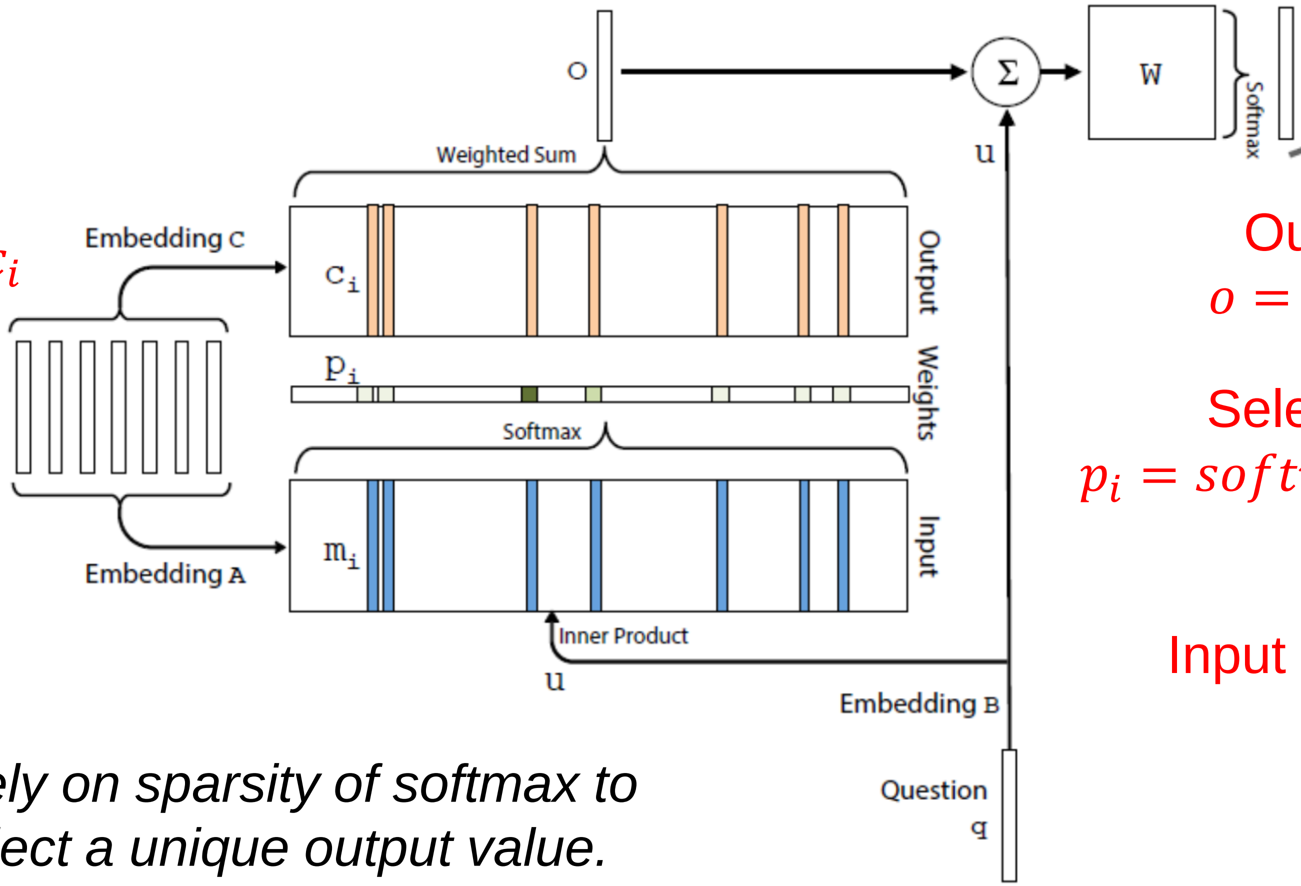


End-to-End Memory Networks

- A closer look at the memory mechanism

Values c_i

Keys m_i



Output
$o = \sum_i p_i c_i$

Selection
$p_i = softmax(u^T m_i)$

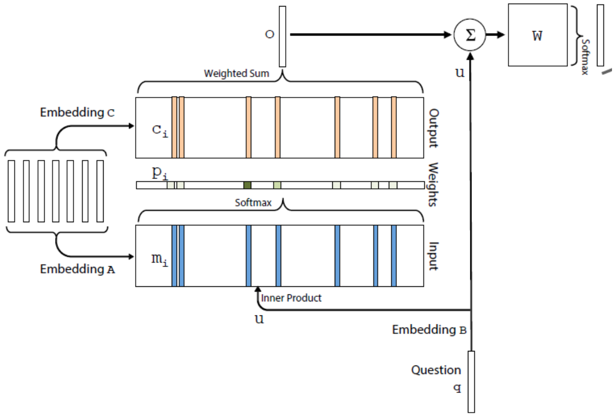
Input query u

$\Rightarrow$ *Rely on sparsity of softmax to select a unique output value.*

S. Sukhbaatar, A. Szlam, J. Weston, R. Fergus, End-to-End Memory Networks. In NIPS 2015.

Image from [Sukhbaatar et al., 2015]

Memory Networks



- Problem with this design
  - Softmax used for the selection involves a normalization over all stored keys.
  - Memory cells that are not accessed get almost zero gradient.
  - When a backpropagation step causes the accessed memory cell to change, this massively affects the gradient flow.

Output
$o = \sum_i p_i c_i$

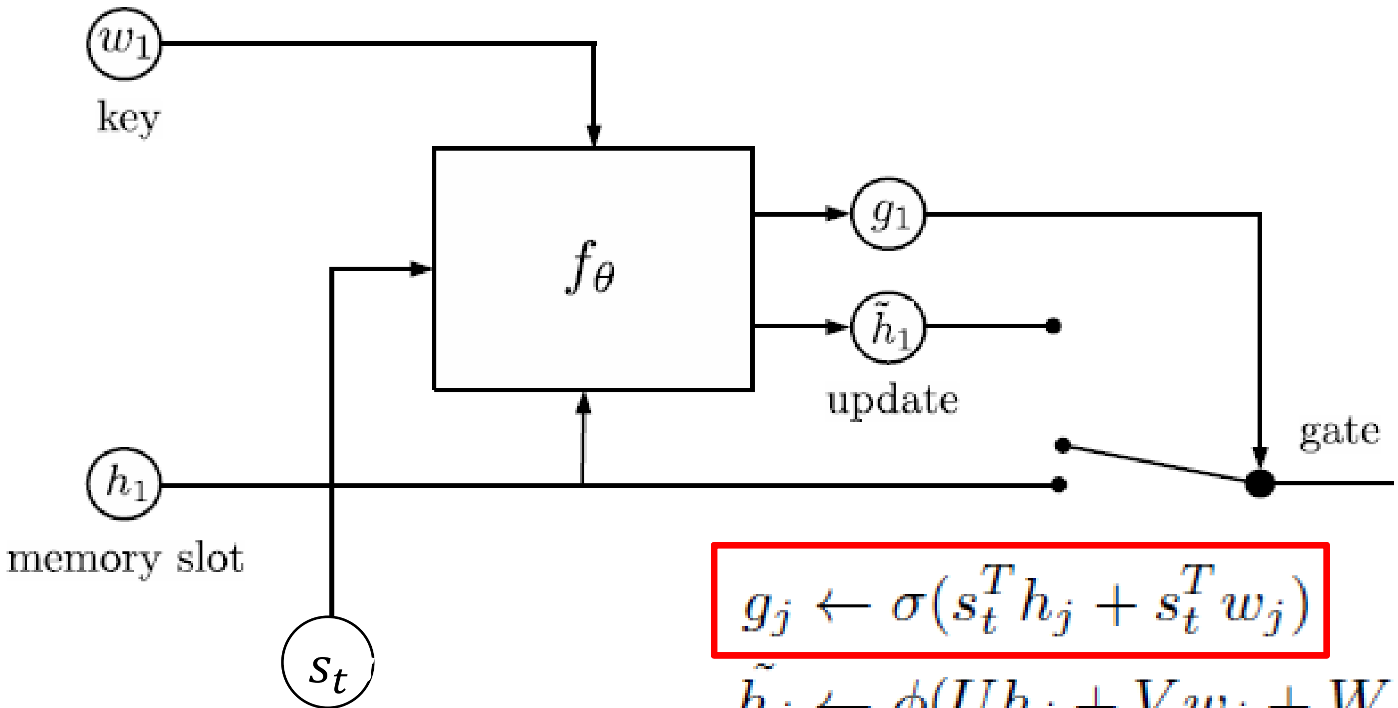
Selection
$p_i = softmax(u^T m_i)$

$\Rightarrow$ Together, this results in bad gradient propagation during learning.

$\Rightarrow$ Very finicky behavior...

Improved Design

- Gated memory (e.g., Recurrent Entity Network)



Gating mechanism

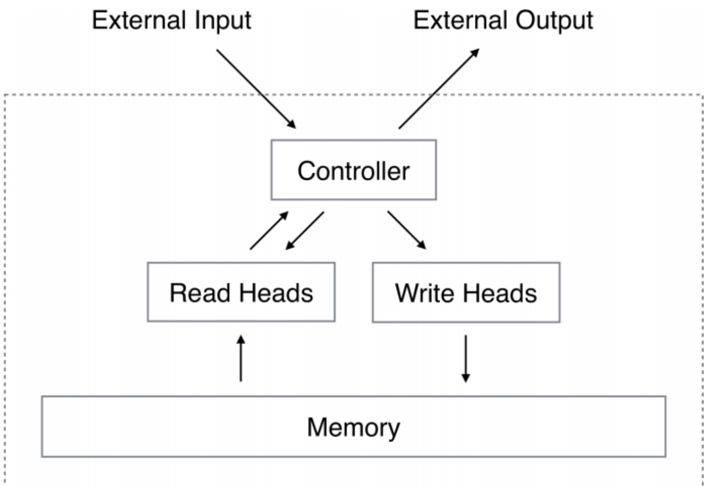
$$\boxed{g_j \leftarrow \sigma(s_t^T h_j + s_t^T w_j)}$$

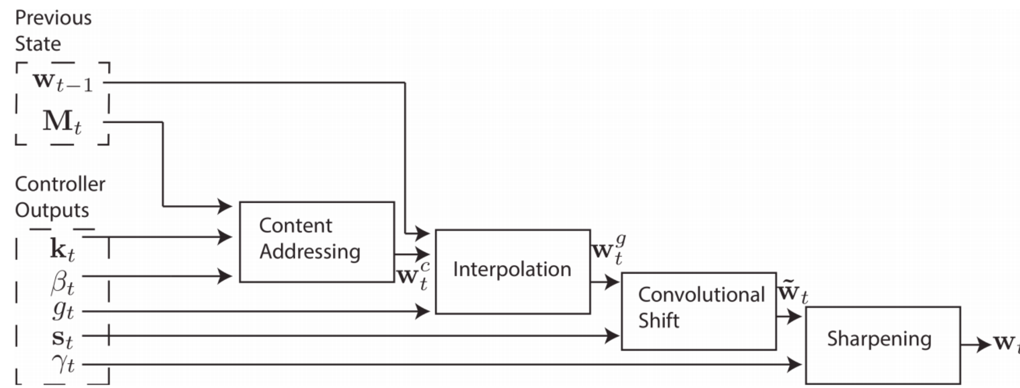
$$\tilde{h}_j \leftarrow \phi(U h_j + V w_j + W s_t)$$

$$h_j \leftarrow h_j + \boxed{g_j \odot \tilde{h}_j} \qquad h_j \leftarrow \frac{h_j}{||h_j||}$$

M. Henaff, J. Weston, A. Szlam, A. Border, Y. LeCun, Tracking the World State with Recurrent Entity Networks. arXiv 1612.03969, 2016.

Neural Turing Machines

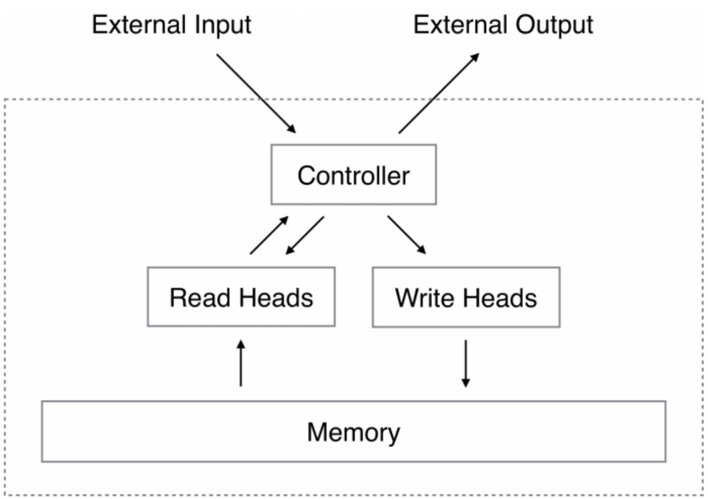


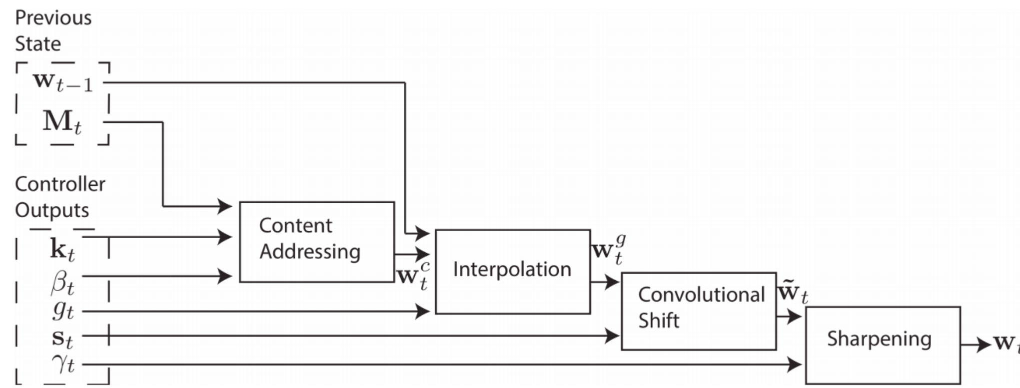


- Goal: Enable general computation with Neural Nets
  - Again key is to make all operations differentiable.
  - Memory + Access operators + Controller
  - Learn entire algorithms from examples.

A. Graves, G. Wayne, I. Danihelka, Neural Turing Machines. arXiv 1410.5401, 2014.

Neural Turing Machines





- Goal: Enable general computation with Neural Nets
  - Again key is to make all operations differentiable.
  - Memory + Access operators + Controller
  - Learn entire algorithms from examples.

A. Graves, G. Wayne, I. Danihelka, Neural Turing Machines. arXiv 1410.5401, 2014.

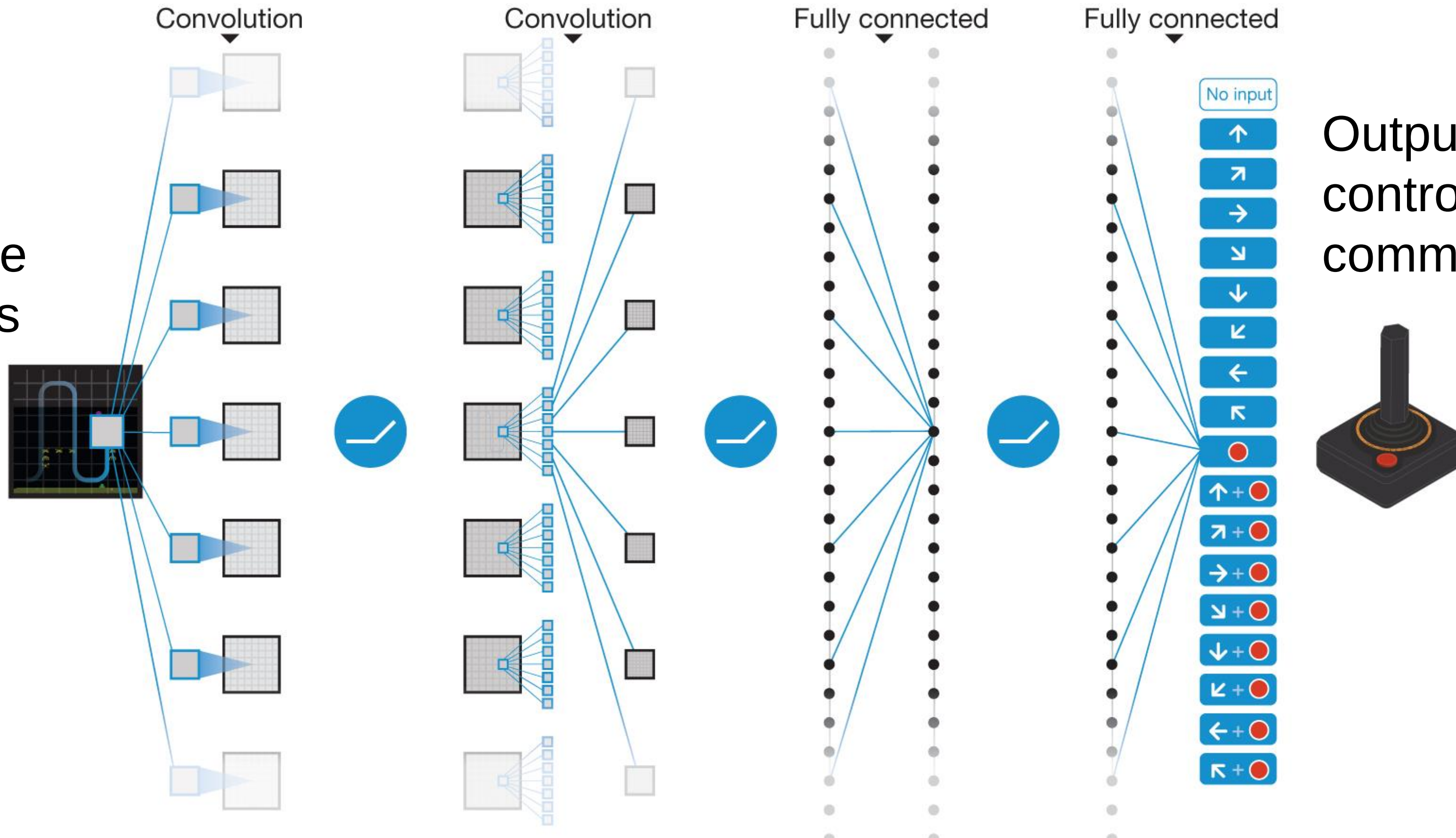
Currently Hot Research Directions

- Generative Models
  - Networks for image generation
  - Generative Adversarial Networks (GAN)
- Towards General Models of Computation
  - Memory Networks
  - Neural Turing Machines
- Deep Reinforcement Learning

Deep Reinforcement Learning

- Example application: Learning to play Atari games

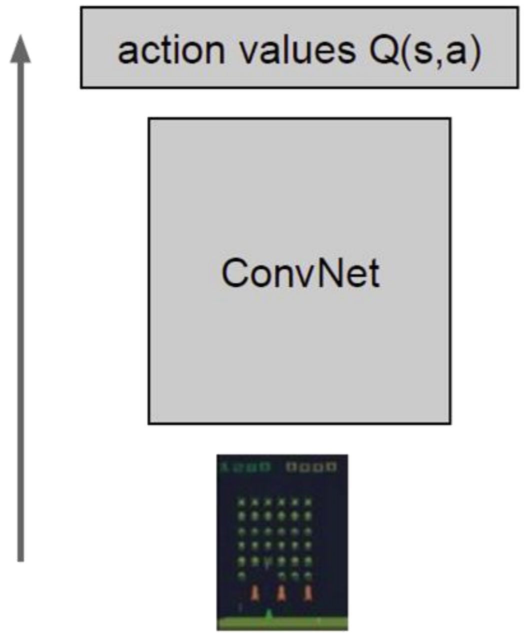
Input: pixels +game scores



Output: control commands

V. Mnih et al., [Human-level control through deep reinforcement learning](), Nature Vol. 518, pp. 529-533, 2015

Idea Behind the Model



- Interpretation
  - Assume finite number of actions
  - Each number here is a real-valued quantity that represents the Q function in Reinforcement Learning

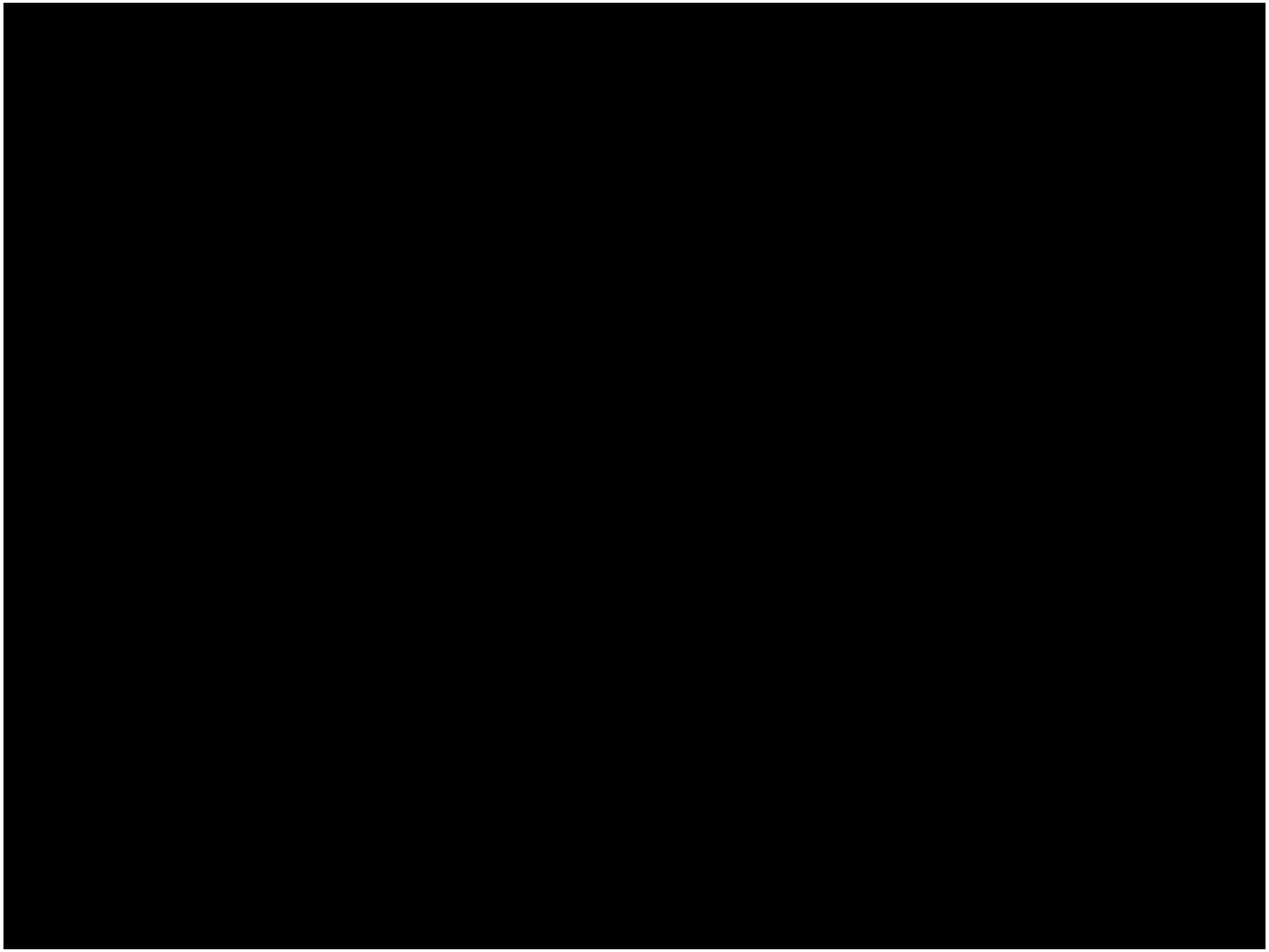
- Collect experience dataset:
  - Set of tuples {(s,a,s',r), … }
  - (State, Action taken, New state, Reward received

- **L2 Regression Loss**

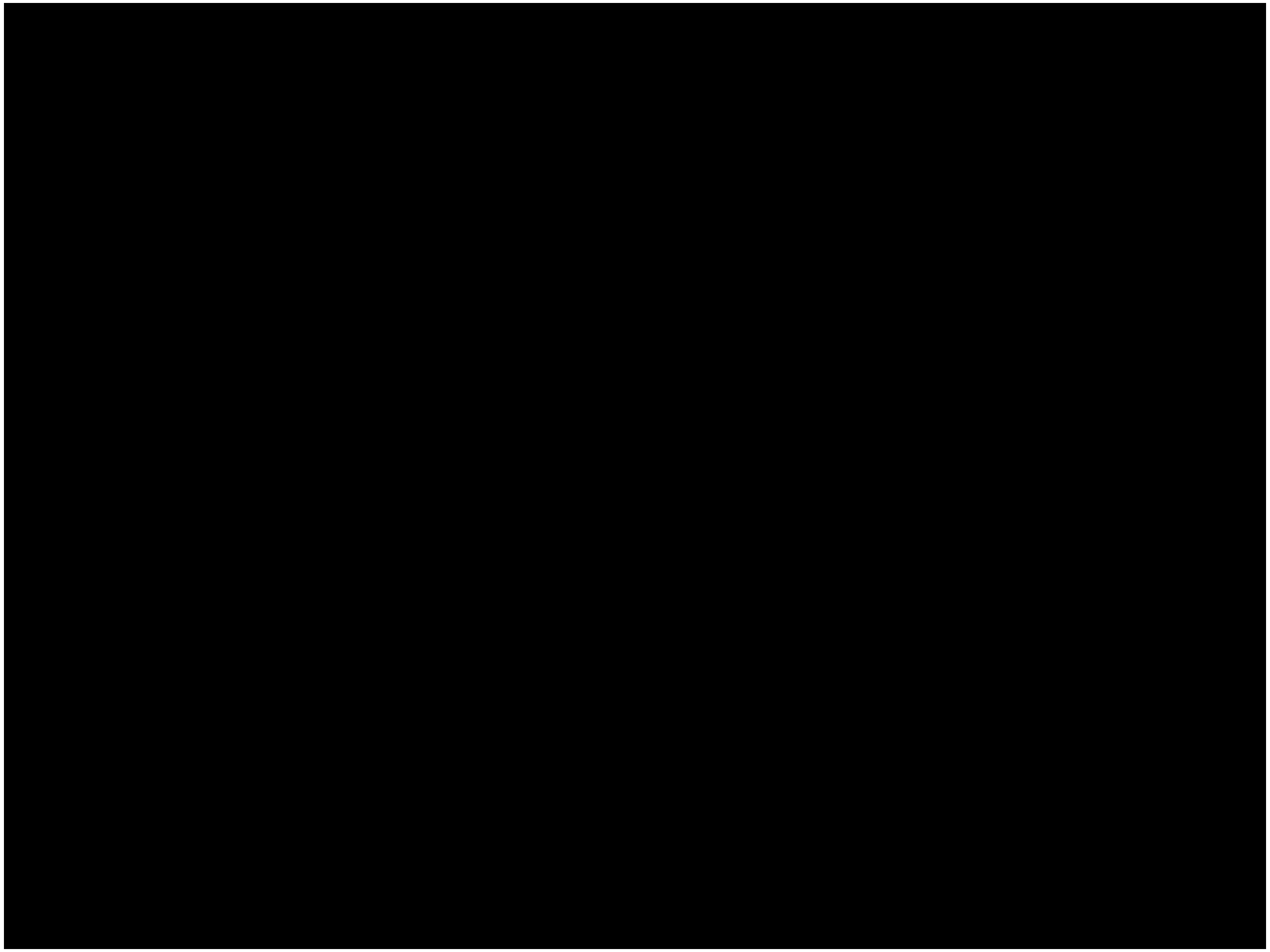
$$L_i(\theta_i) = \mathbb{E}_{(s,a,r,s')\sim \mathrm{U}(D)}\left[\left(\overset{\textbf{target value}}{\boxed{r+\gamma \max_{a'} Q(s',a';\theta_i^-)}} - \overset{\textbf{predicted value}}{\boxed{Q(s,a;\theta_i)}}\right)^2\right]$$

Current reward + estimate of future reward, discounted by γ

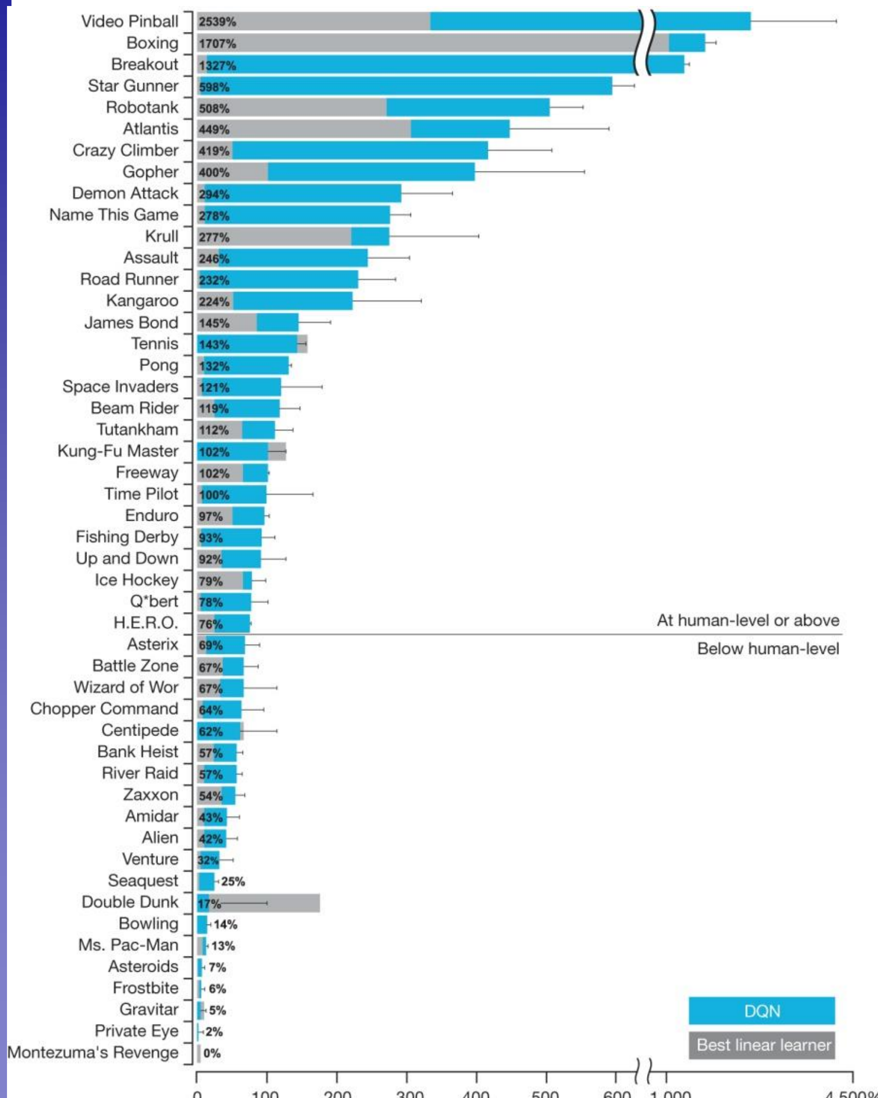
Results: Space Invaders

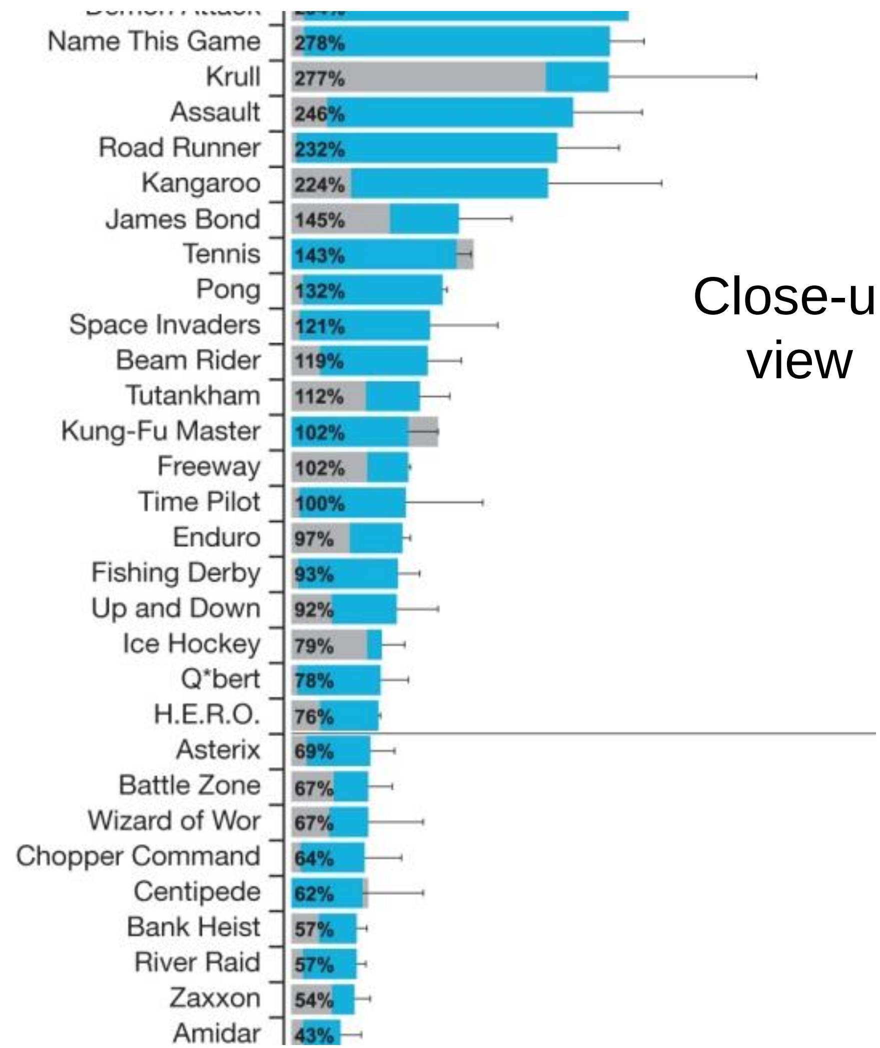


Results: Breakout



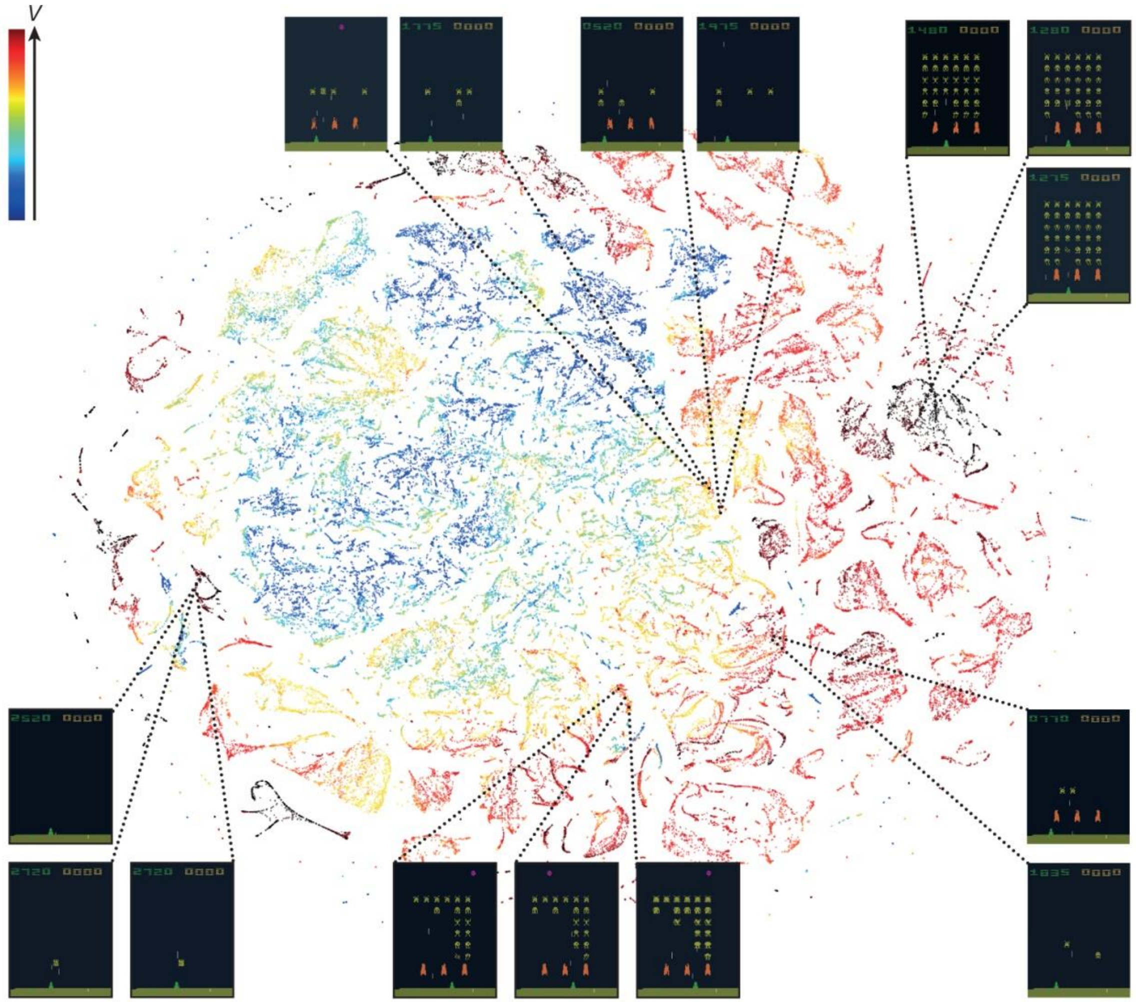
Comparison with Human Performance





Close-up view

Learned Representation



- t-SNE embedding of DQN last hidden layer (Space Inv.)

Success Story: Alpha Go

References and Further Reading

- Generative Adversarial Networks (GANs)
  - I.J. Goodfellow,J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Networks, arXiv:1406.2661, 2014.
  - M. Arjovsky, S. Chintala, L. Boutou, Wasserstein GAN, arXiv:1701.07875, 2017.
  - L. Mescheder, P. Gehler, A. Geiger, The Numerics of GANs, arXiv:1705.10461, 2017.

References and Further Reading

- Memory Networks
    - S. Sukhbaatar, A. Szlam, J. Weston, R. Fergus, End-to-End Memory Networks. In NIPS 2015.
    - M. Henaff, J. Weston, A. Szlam, A. Border, Y. LeCun, Tracking the World State with Recurrent Entity Networks. arXiv 1612.03969, 2016.

- Neural Turing Machines
    - A. Graves, G. Wayne, I. Danihelka, Neural Turing Machines. arXiv 1410.5401, 2014.

References and Further Reading

- DQN paper
  - www.nature.com/articles/nature14236

- AlphaGo paper
  - www.nature.com/articles/nature16961