What’s in a good representation?

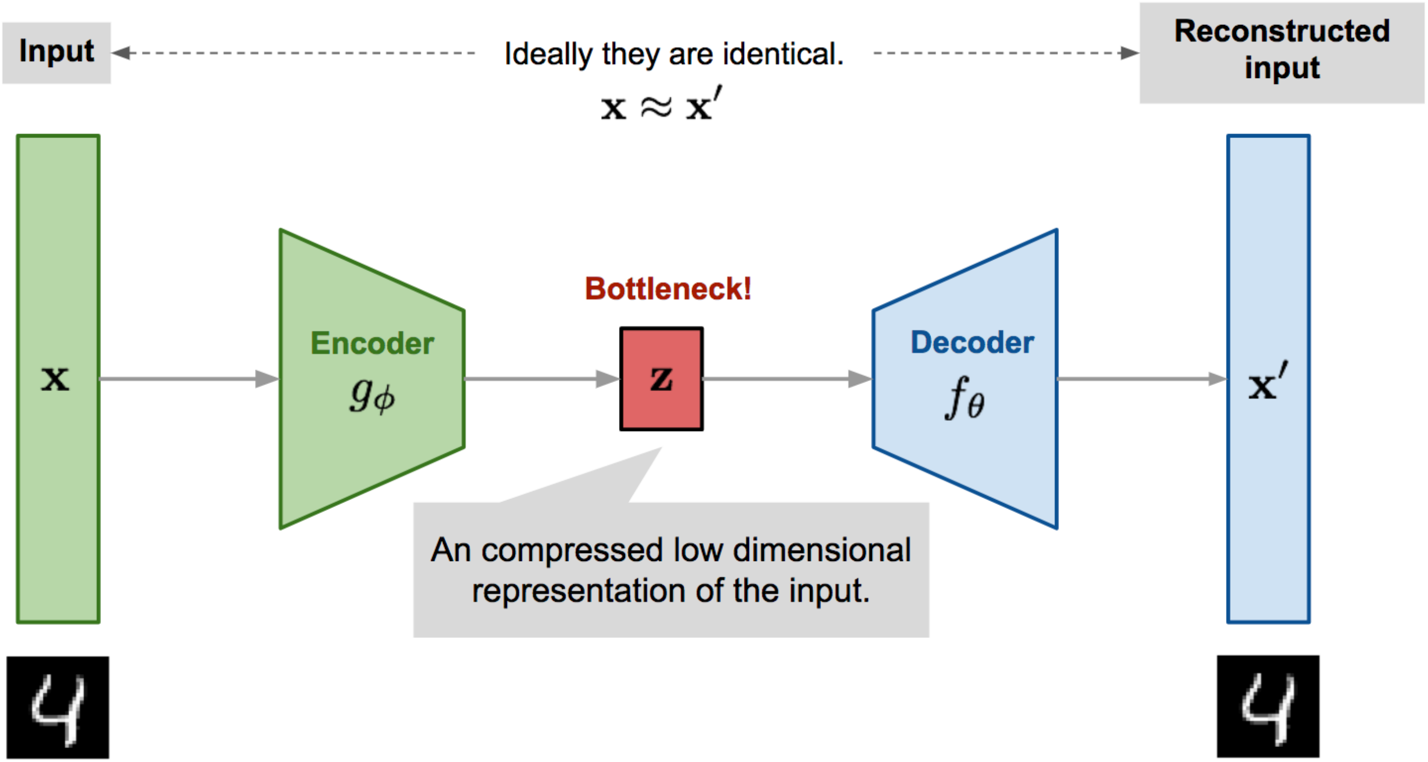
John Tan Chong Min

Aim

- Find a way to encode a suitable representation to perform decision making
- Such representation can also be how we store memories for use in the future

Autoencoders: Representation via Reconstruction

- Prioritises output clarity – may not disentangle well in latent space


Input
Ideally they are identical.
$\mathbf{x} \approx \mathbf{x}'$
Reconstructed input
$\mathbf{x}$
Encoder
g_ϕ
Bottleneck!
$\mathbf{z}$
Decoder
f_θ
$\mathbf{x}'$
An compressed low dimensional representation of the input.

Do you need to predict everything?

Transformers: Representation via Prediction

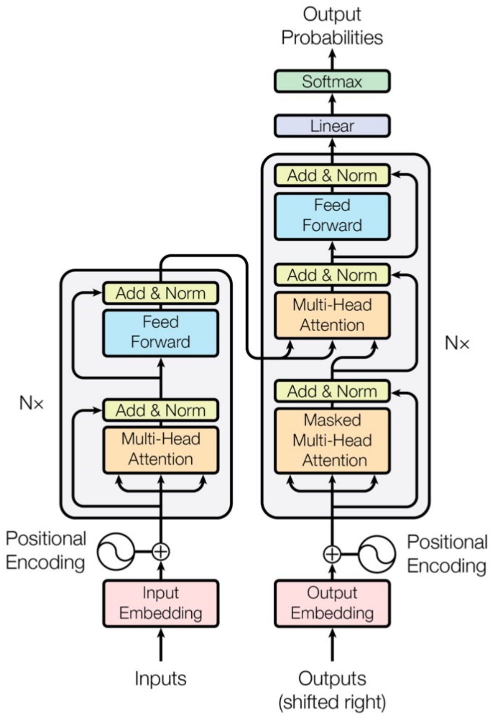
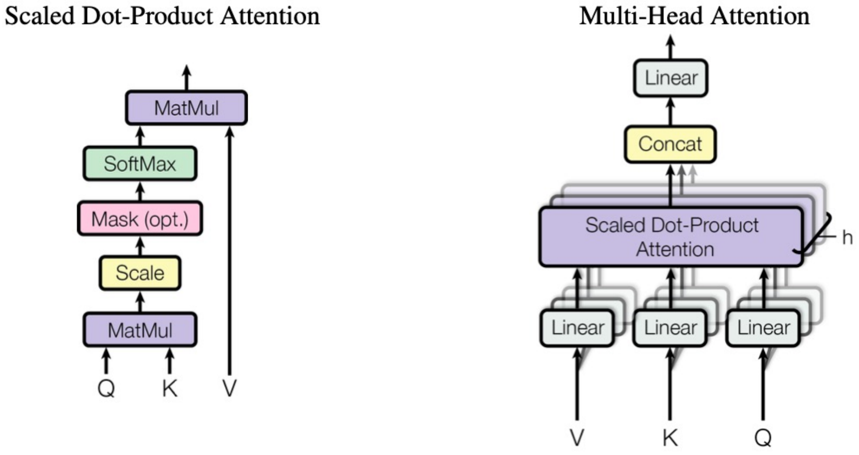


Figure 1: The Transformer - model architecture.


Scaled Dot-Product Attention

Multi-Head Attention

Figure 2: (left) Scaled Dot-Product Attention. (right) Multi-Head Attention consists of several attention layers running in parallel.

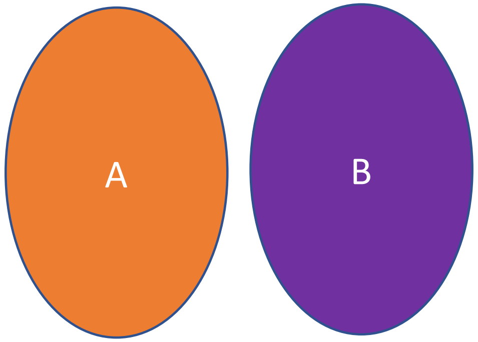
$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

Taken from: Attention is all you need. Vaswani et al. (2017)

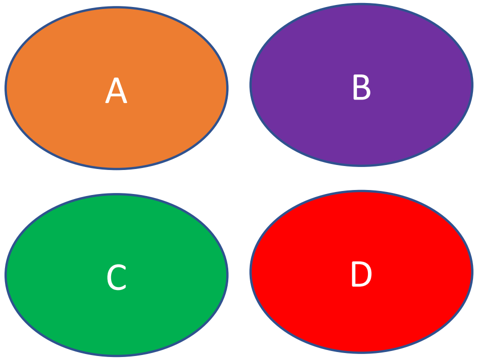
Large Self-Supervised Learning

- Self-supervised learning helps to learn better manifolds across large data
- Can work zero-shot on a new sample

Manifold of having only 2 classes

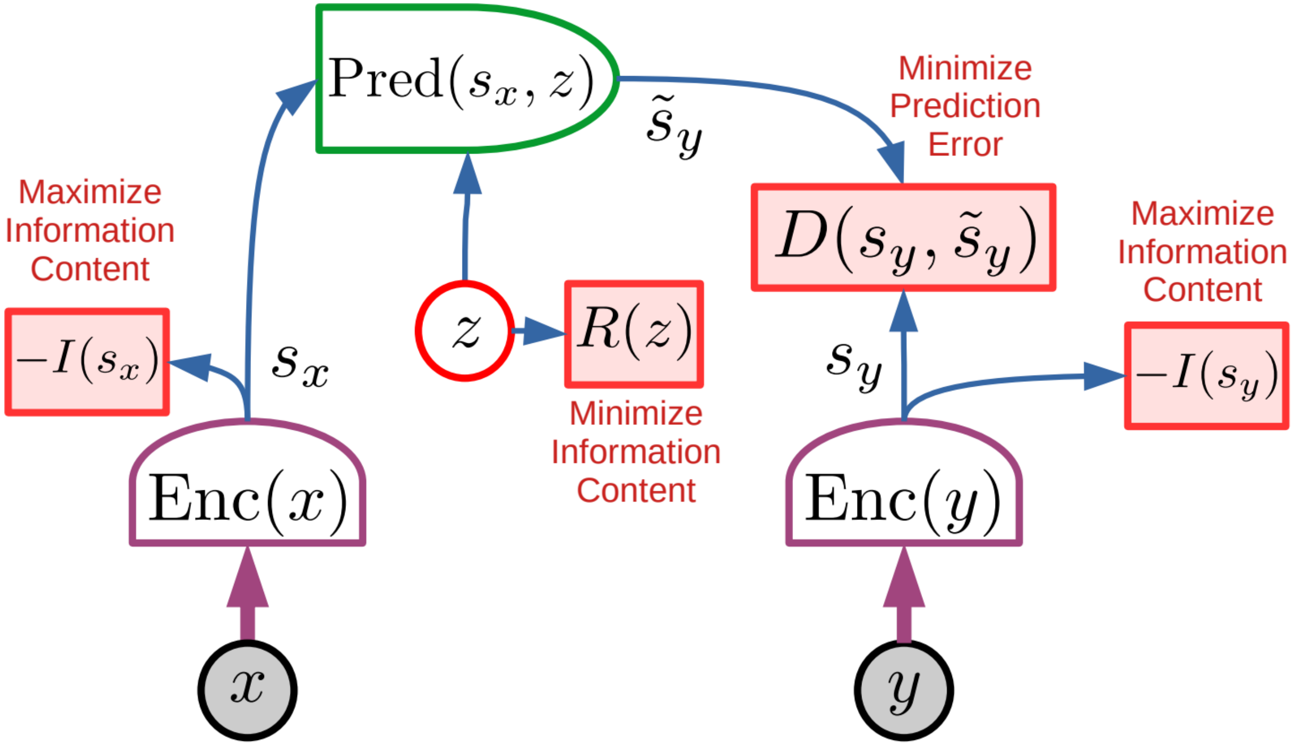


Manifold of having 4 classes



JEPA - Only use whatever is necessary to predict

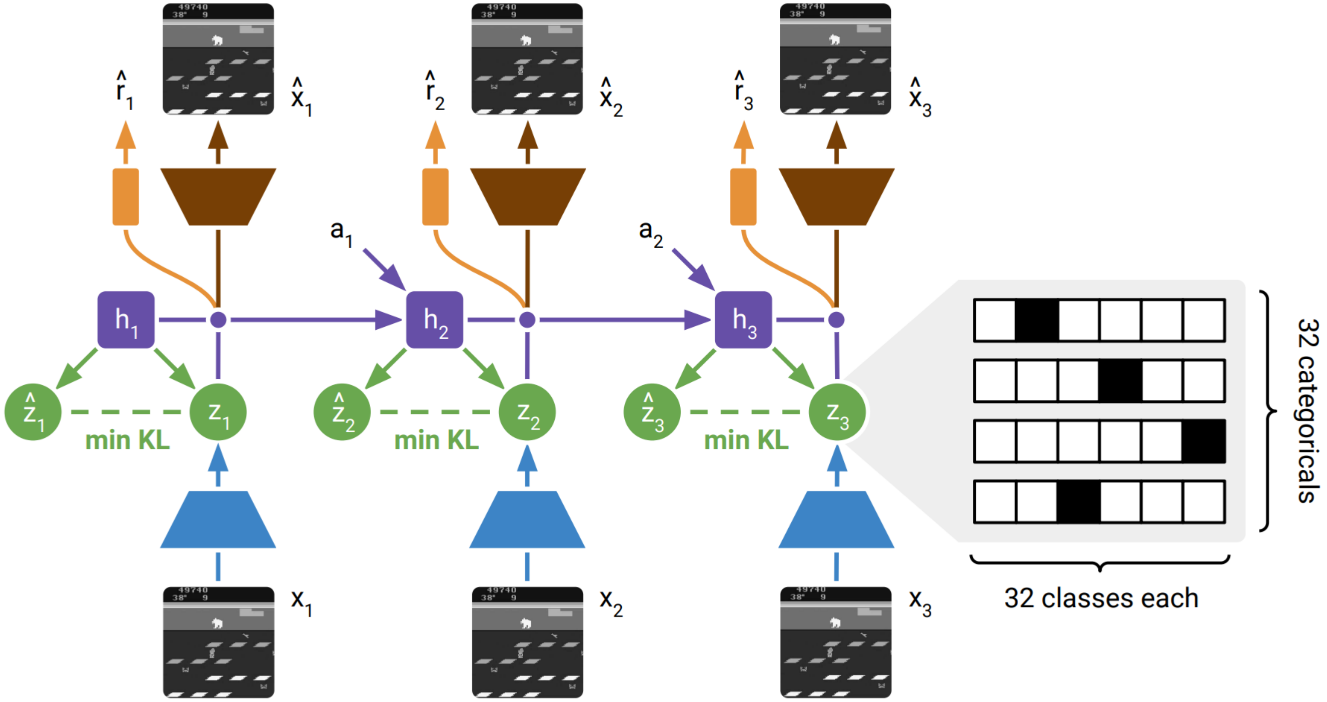
- Prediction is done in latent space



A Path towards Autonomous Machine Intelligence. Yann LeCun. 2022.

World Modelling

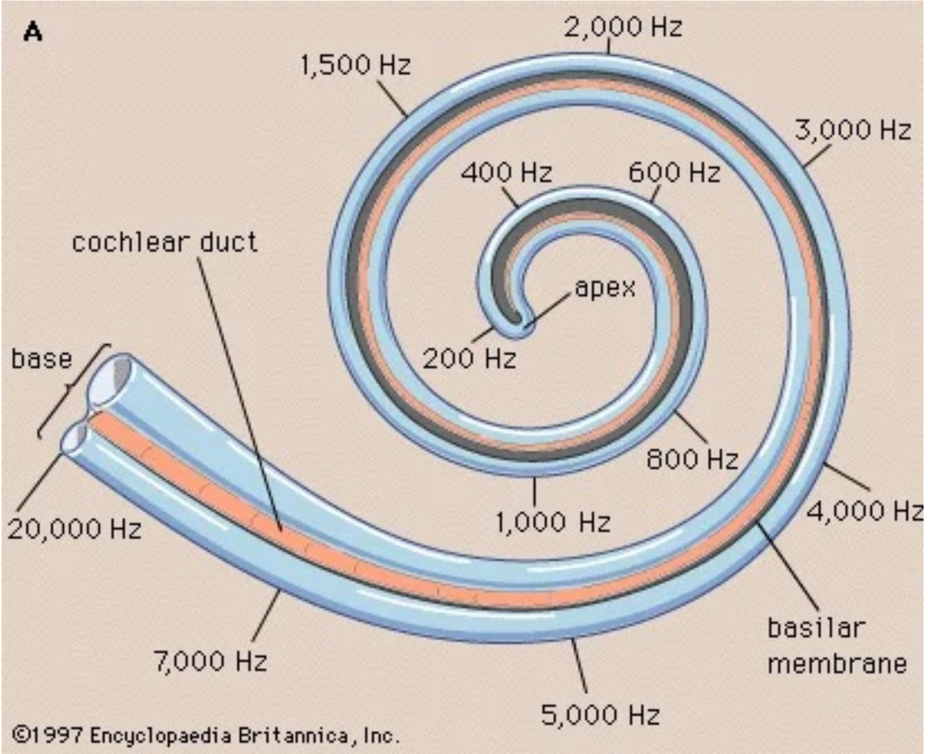
- Use hidden representations for prediction
- Latent space is **discrete** using latent space of categorical variables



Mastering Atari with Discrete World Models. Hafner et al. 2022.

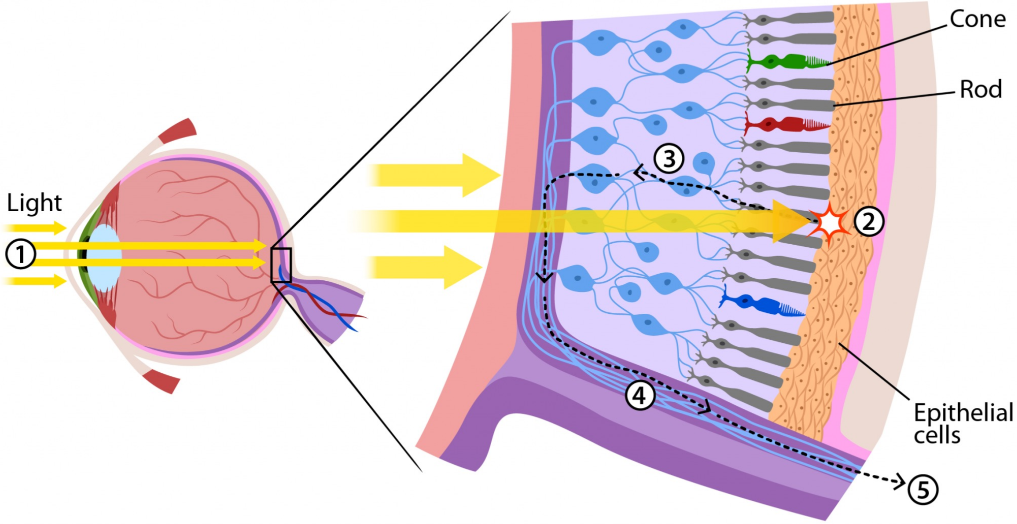
Natural Fixed Biases: Faster learning by constraints

- Sound: frequency in cilia

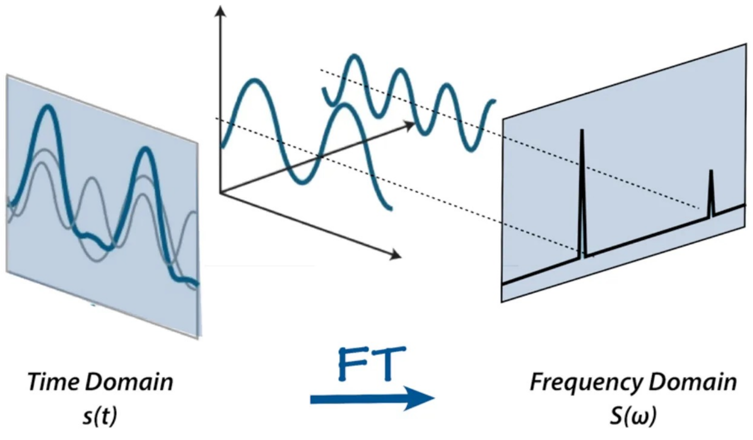


©1997 Encyclopaedia Britannica, Inc.

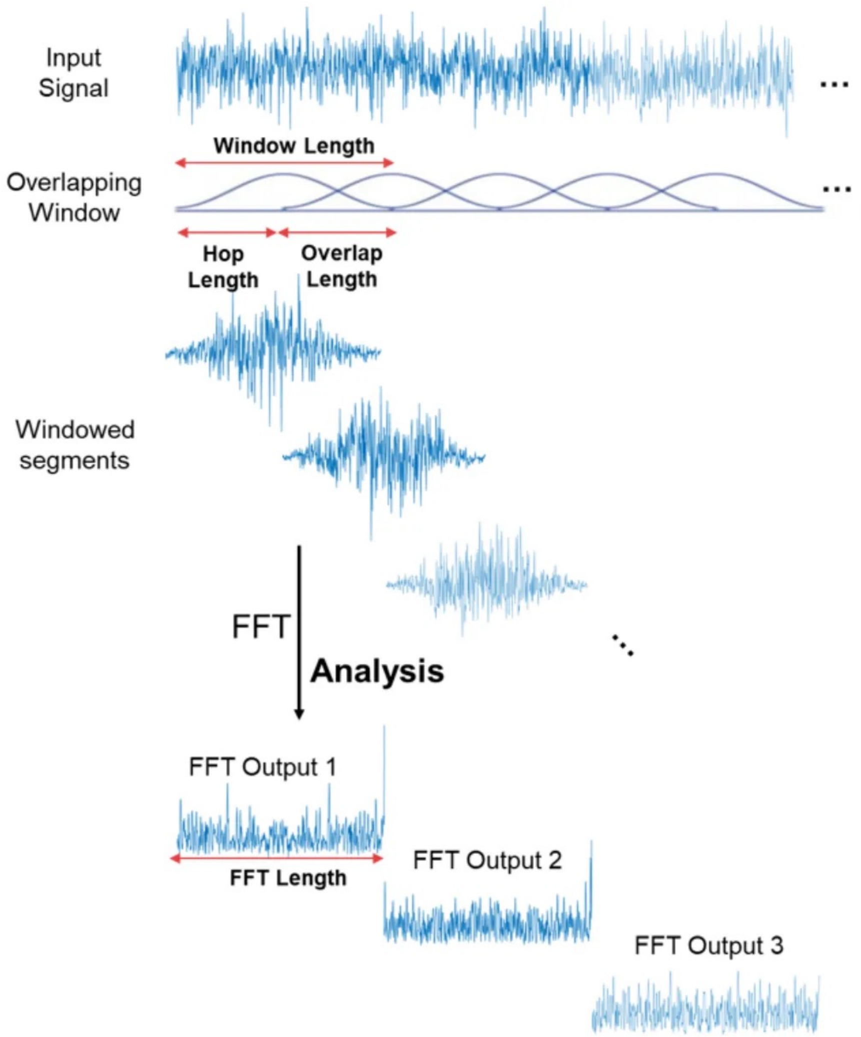
- Vision:
  - Local patches
  - Cones for Red, Green, Blue
  - Rods for Black and White



Audio – Freq Modelling



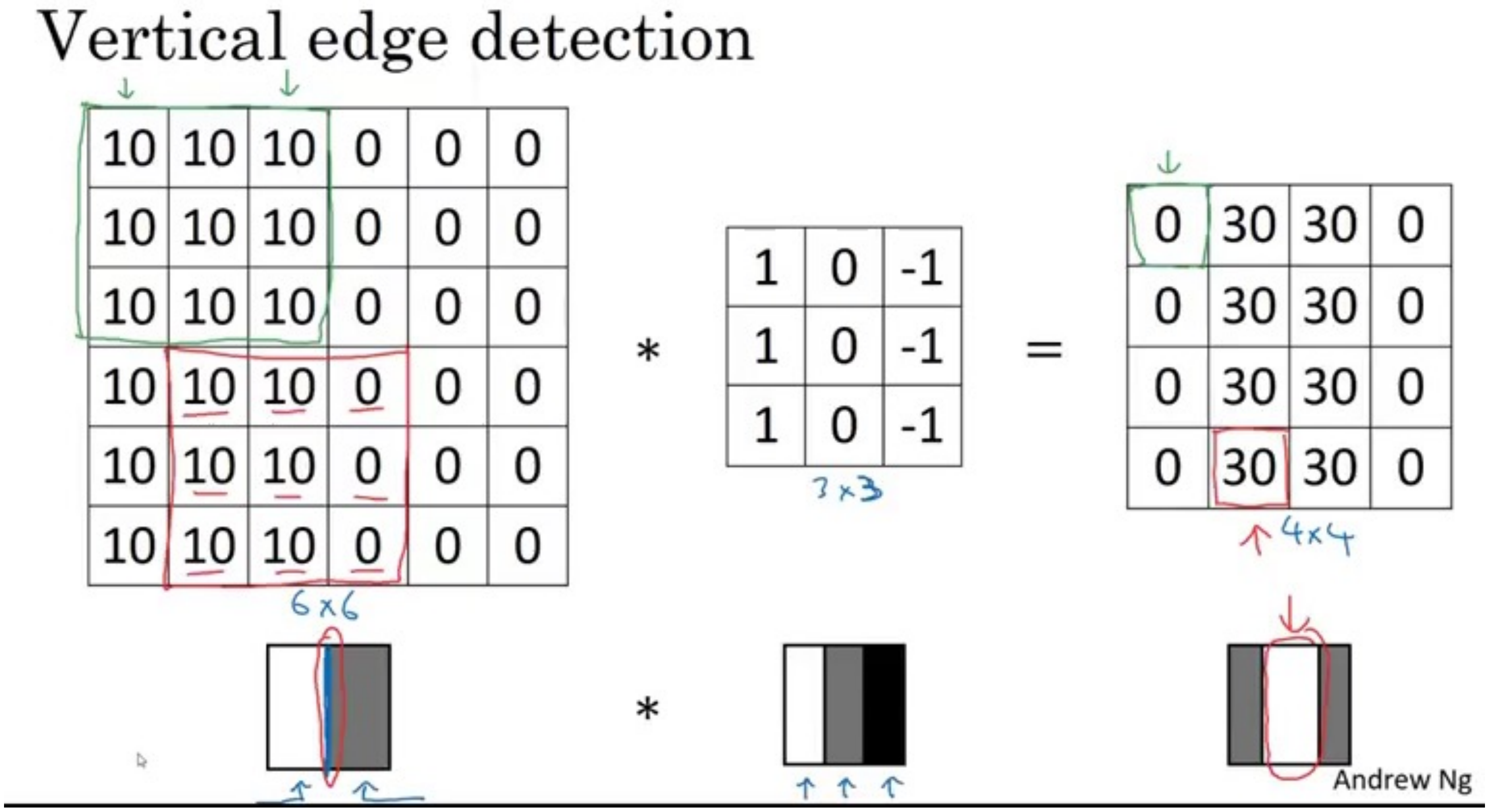
Audio signal converted into frequencies



Use overlapping window to model waveform over time

Images taken from: https://medium.com/analytics-vidhya/understanding-the-mel-spectrogram-fca2afa2ce53

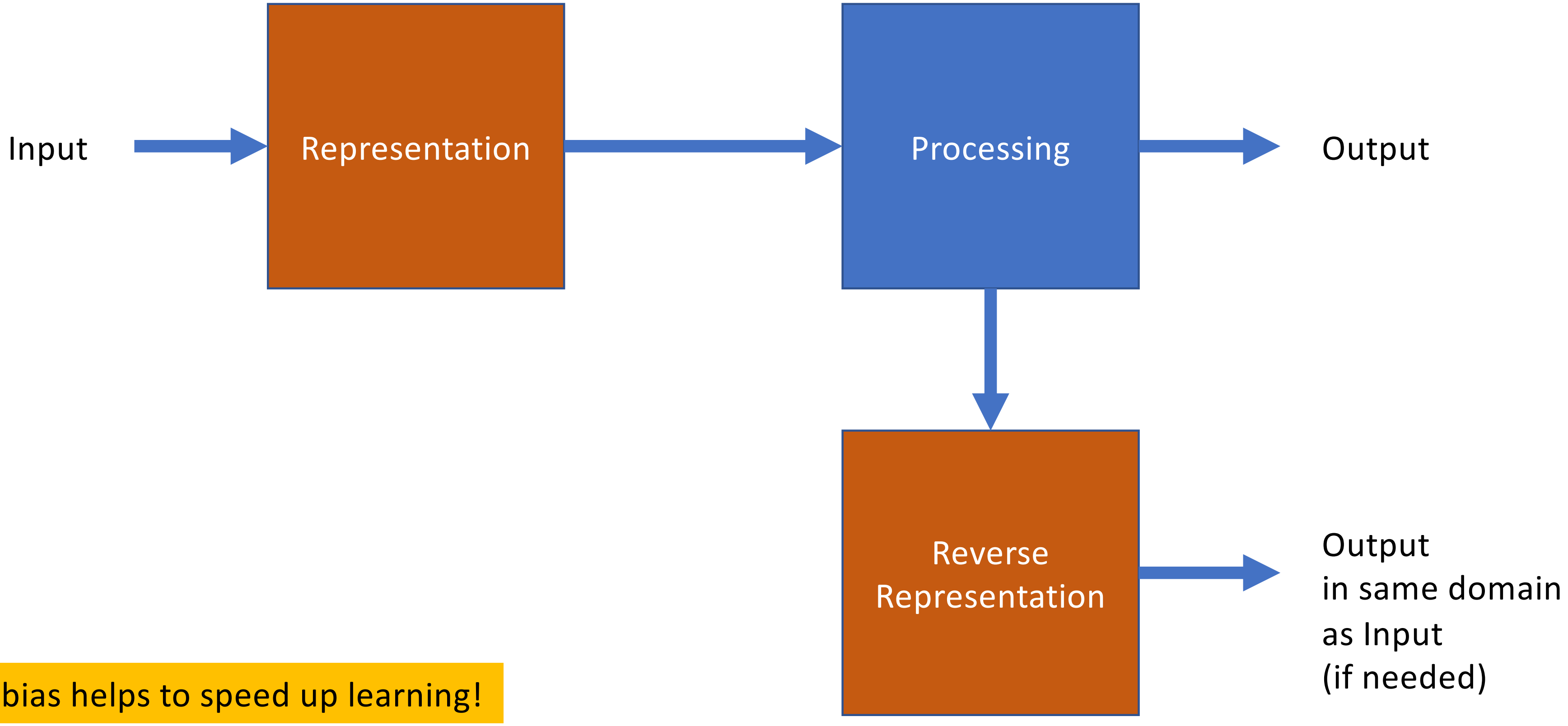
Vision – Pixel Proximity



Vertical edge detection

$$\underbrace{\begin{bmatrix} 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \end{bmatrix}}_{6\times 6} * \underbrace{\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}}_{3\times 3} = \underbrace{\begin{bmatrix} 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \end{bmatrix}}_{4\times 4}$$

Information Pipeline – Bias for Representation



Input

Representation

Processing

Output

Reverse Representation

Output in same domain as Input (if needed)

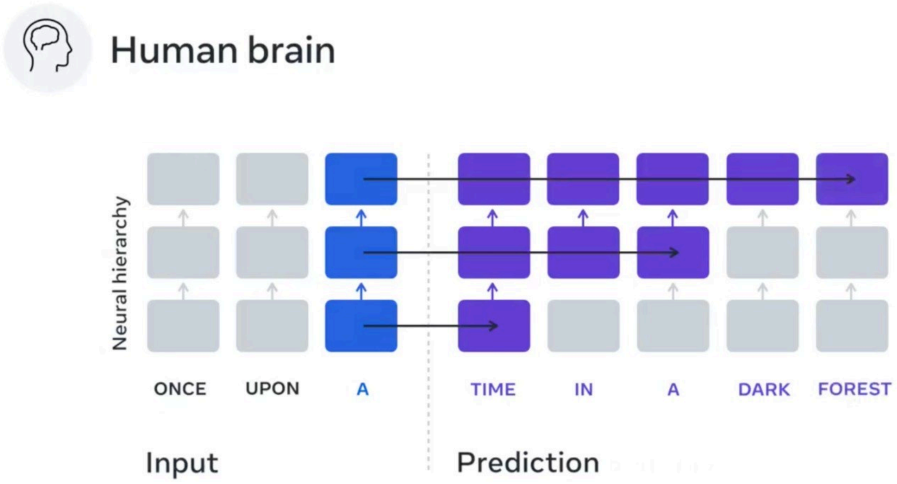
Fixed bias helps to speed up learning!

The next level

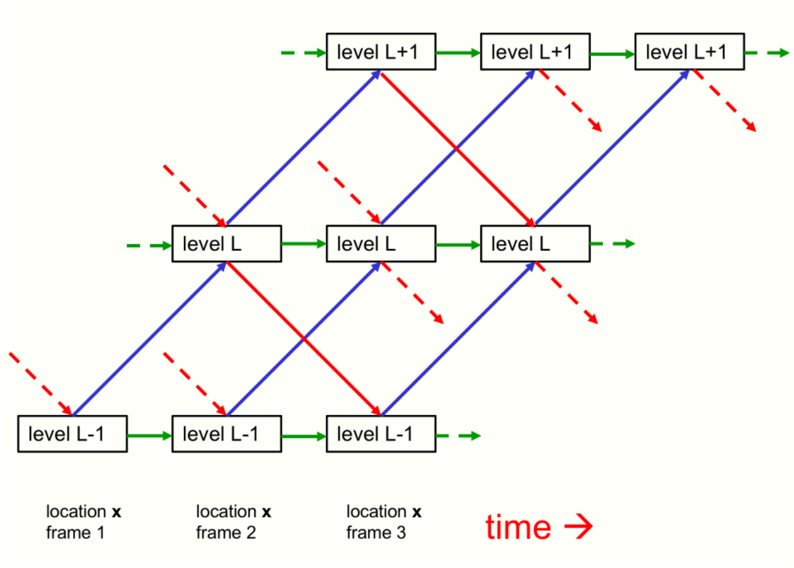
Hierarchical Prediction

Hierarchical Prediction is the future

- Hierarchical prediction of more than just next token, but broader prediction at higher levels
- Higher level prediction can be more abstract and less detailed than lower levels



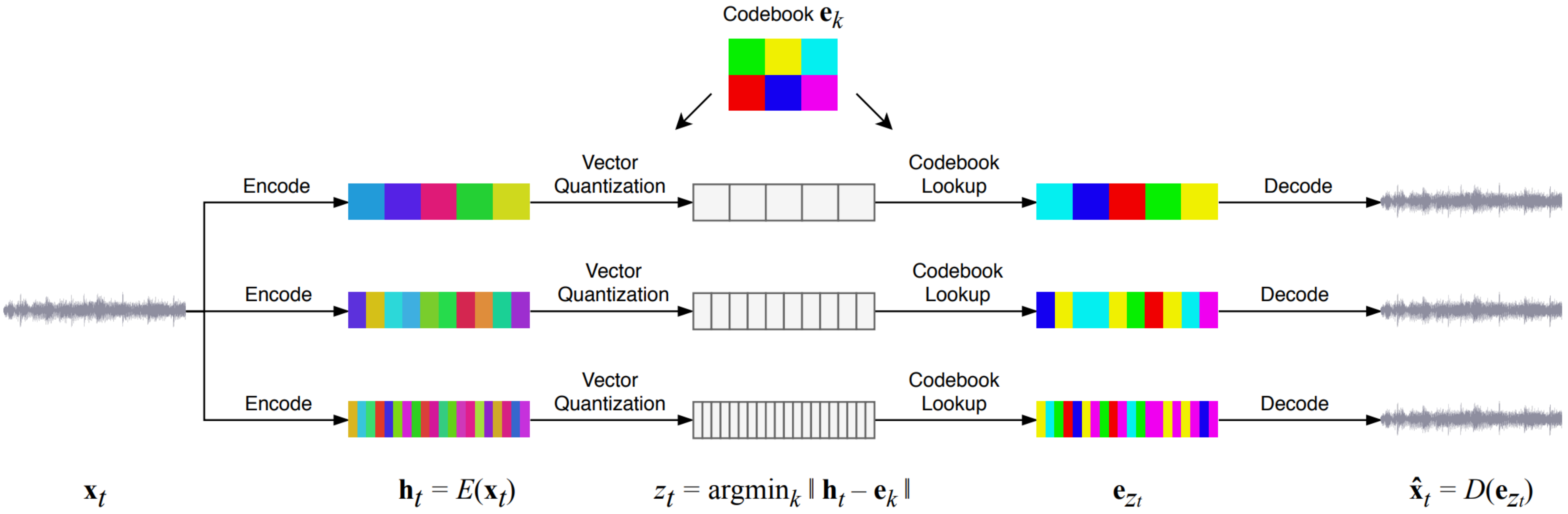
Evidence of a predictive coding hierarchy in the human brain listening to speech. Caucheteux. 2022. Nature Human Behaviour.



How to represent part-whole hierarchies in a neural network. Hinton. 2021.

Hierarchical Prediction – Jukebox (OpenAI)

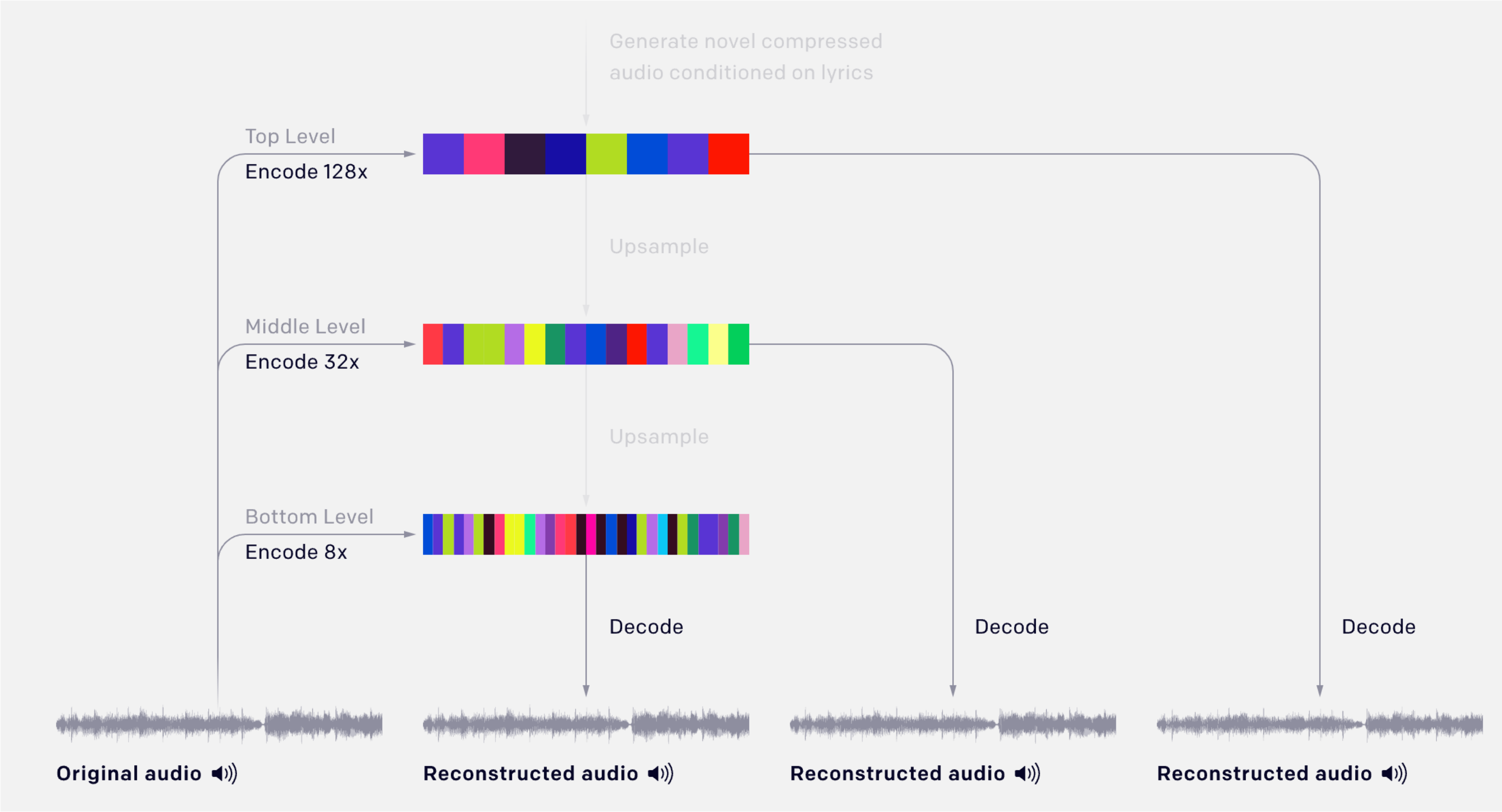
- Hierarchical Quantization of sound from coarse-grained input to fine-grained layers
- Codebook entries are **finite**



Jukebox: A Generative Model for Music. Dhariwal et. al. 2020.

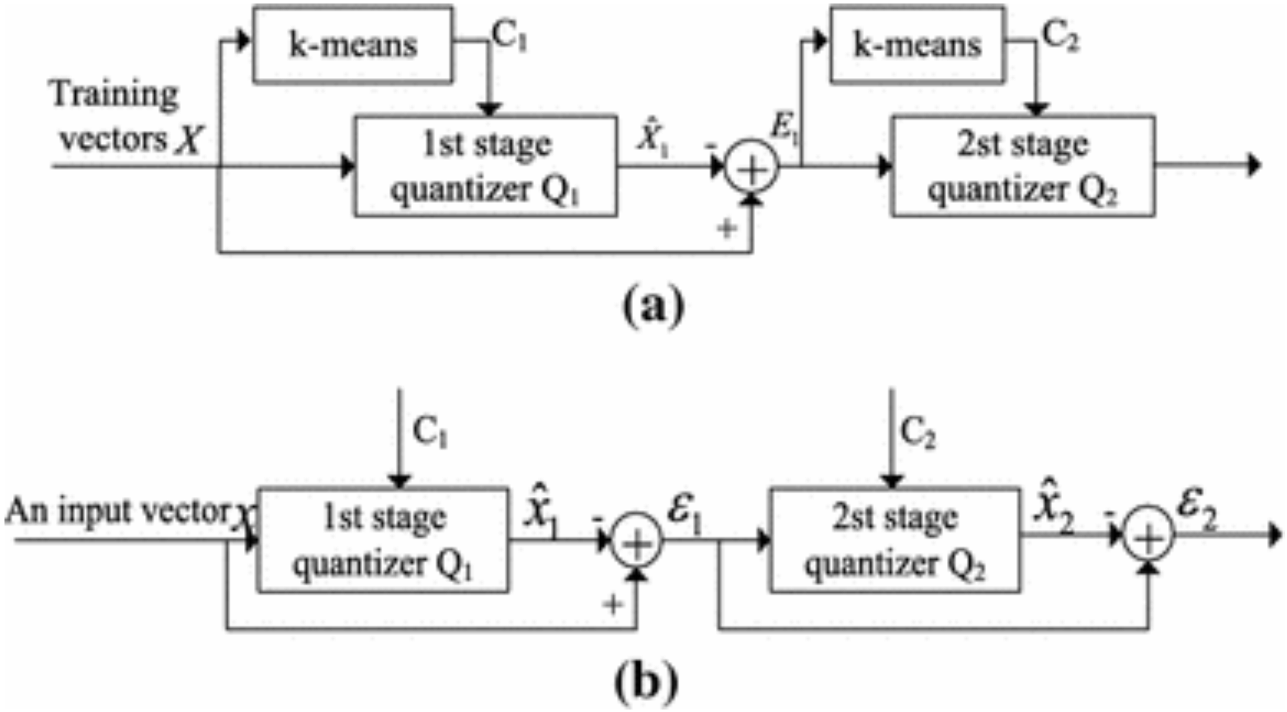
Hierarchical Prediction – Jukebox (OpenAI)

- Conditional generation of coarse-grained input to fine-grained layers



Residual Vector Quantization

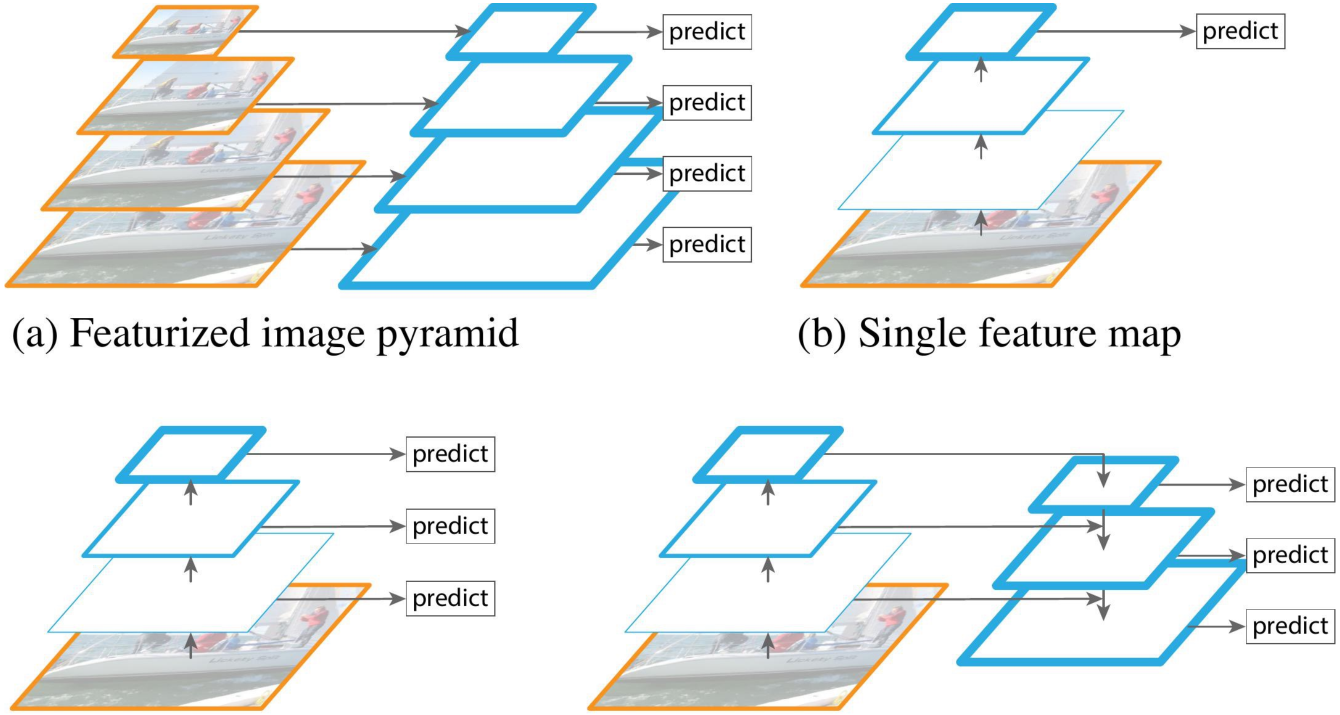
- Hierarchical Quantization from coarse-grained input to fine-grained layers
- Akin to compositionality?



Optimized residual vector quantization for efficient approximate nearest neighbor search. Ai et al. 2015.

Hierarchical Prediction - Feature Pyramid Network

- Hierarchical prediction from coarse-grained image to fine-grained image



(a) Featurized image pyramid

(b) Single feature map

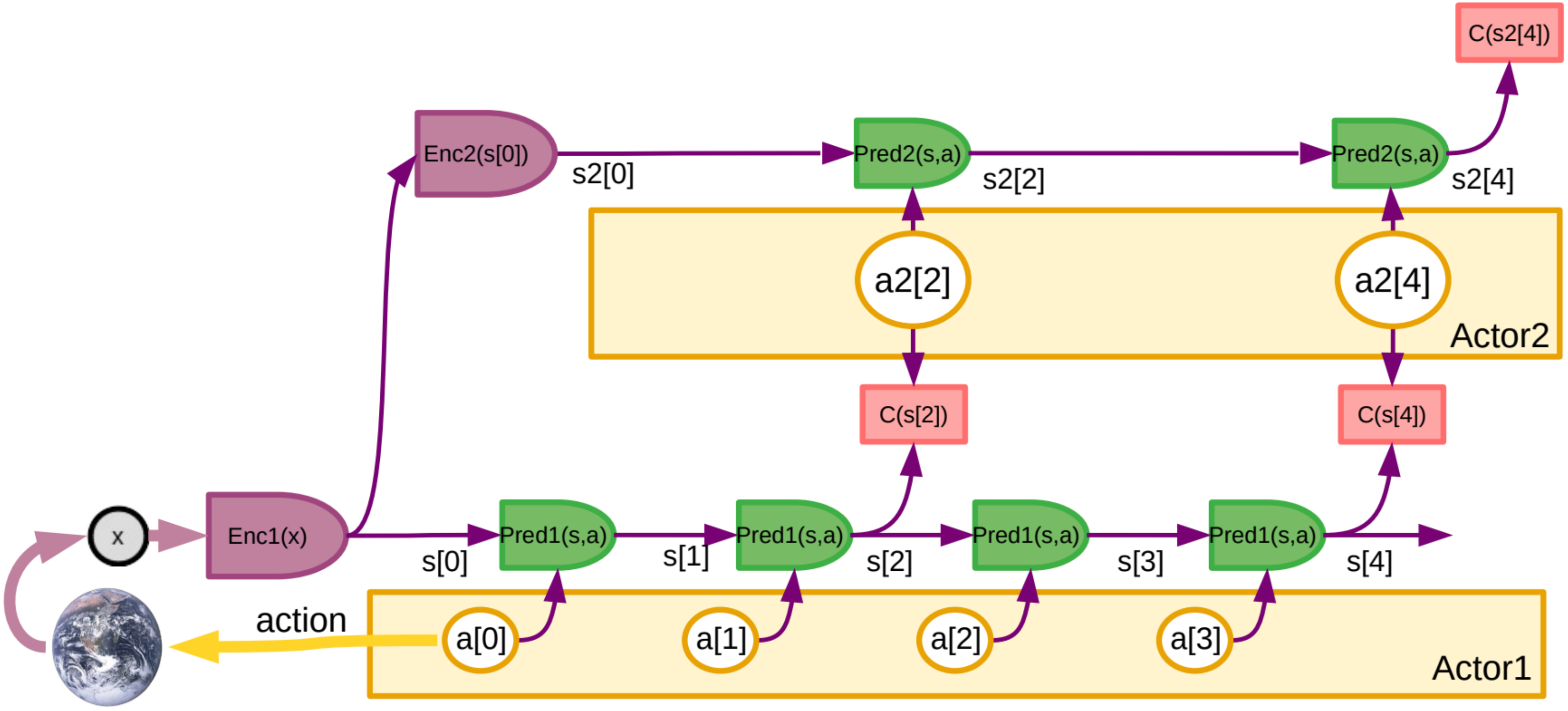
(c) Pyramidal feature hierarchy

(d) Feature Pyramid Network

Feature Pyramid Networks for Object Detection. Lin et al. 2017.

Hierarchical JEPA

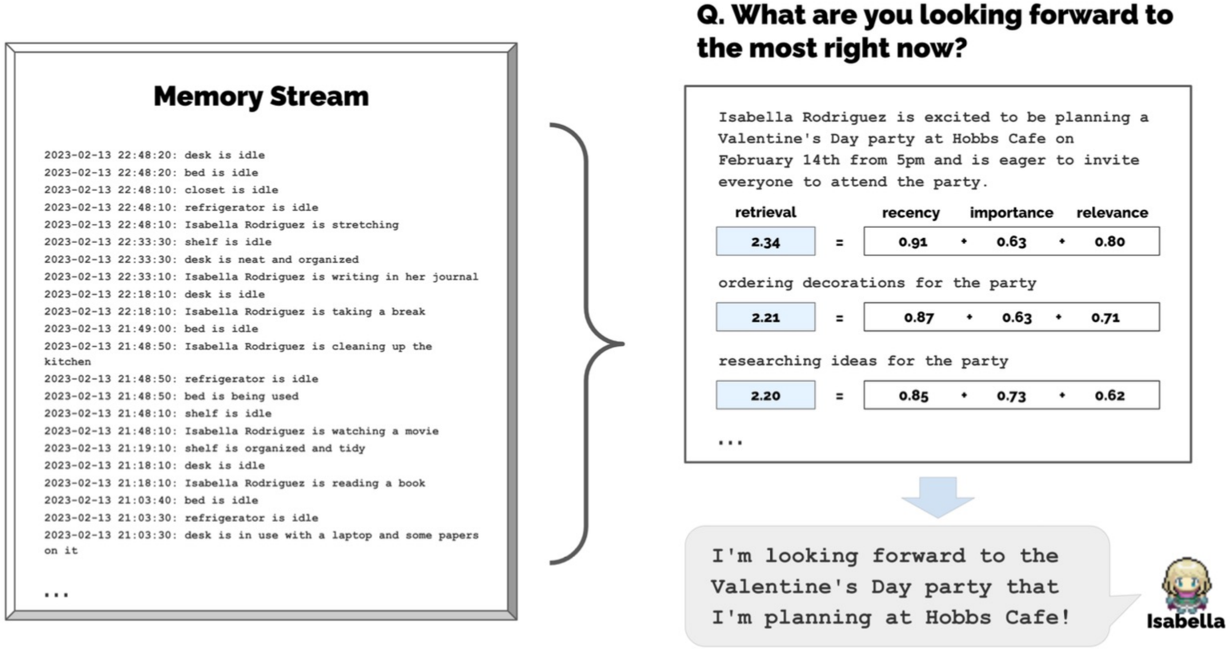
- Hierarchical prediction of actions from the highest level action to the lowest level action



A Path towards Autonomous Machine Intelligence. Yann LeCun. 2022.

Hierarchical Action Prediction

- Hierarchical prompting of actions from broad action to specific actions



```
2023-02-13 22:48:20: desk is idle
2023-02-13 22:48:20: bed is idle
2023-02-13 22:48:10: closet is idle
2023-02-13 22:48:10: refrigerator is idle
2023-02-13 22:48:10: Isabella Rodriguez is stretching
2023-02-13 22:33:30: shelf is idle
2023-02-13 22:33:30: desk is neat and organized
2023-02-13 22:33:10: Isabella Rodriguez is writing in her journal
2023-02-13 22:18:10: desk is idle
2023-02-13 22:18:10: Isabella Rodriguez is taking a break
2023-02-13 21:49:00: bed is idle
2023-02-13 21:48:50: Isabella Rodriguez is cleaning up the
kitchen
2023-02-13 21:48:50: refrigerator is idle
2023-02-13 21:48:50: bed is being used
2023-02-13 21:48:10: shelf is idle
2023-02-13 21:48:10: Isabella Rodriguez is watching a movie
2023-02-13 21:19:10: shelf is organized and tidy
2023-02-13 21:18:10: desk is idle
2023-02-13 21:18:10: Isabella Rodriguez is reading a book
2023-02-13 21:03:40: bed is idle
2023-02-13 21:03:30: refrigerator is idle
2023-02-13 21:03:30: desk is in use with a laptop and some papers
on it

...
```

Q. What are you looking forward to the most right now?

Isabella Rodriguez is excited to be planning a Valentine's Day party at Hobbs Cafe on February 14th from 5pm and is eager to invite everyone to attend the party.

retrieval		recency		importance		relevance
2.34	=	0.91	+	0.63	+	0.80

ordering decorations for the party

2.21	=	0.87	+	0.63	+	0.71

researching ideas for the party

2.20	=	0.85	+	0.73	+	0.62

...

I'm looking forward to the Valentine's Day party that I'm planning at Hobbs Cafe!

Figure 6: The memory stream comprises a large number of observations that are relevant and irrelevant to the agent's current situation. Retrieval identifies a subset of these observations that should be passed to the language model to condition its response to the situation.

Generative Agents: Interactive Simulacra of Human Behavior. Joon et al. 2022

Transformers:
Can a Transformer perform hierarchical generation?

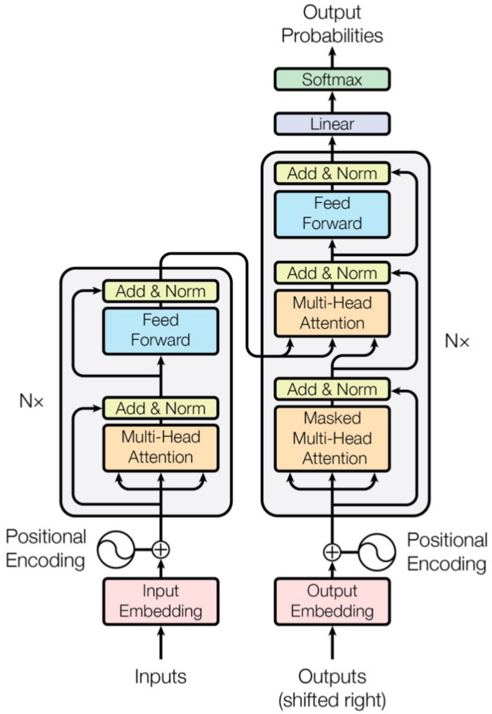
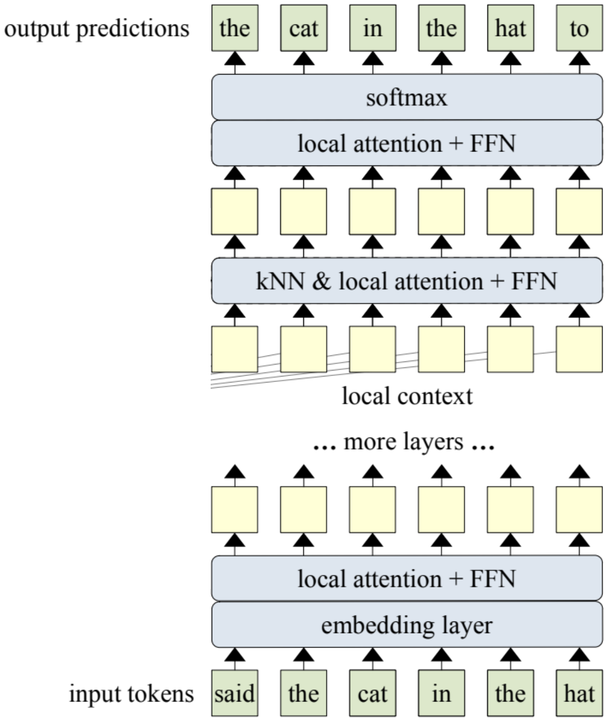


Figure 1: The Transformer - model architecture.

Attention is all you need. Vaswani et al. (2017)



Memorizing Transformers. Wu et al. 2022.

Questions to Ponder

- Do we learn from experience, or from natural fixed bias? Or both?
- Should we do prediction whereby we map back to input space (like tokens in Transformers), or should we just predict the latent space? What are the benefits and drawbacks?
- Should we use hierarchical generation? Is our brain hierarchical or more flat like what Jeff Hawkins proposes in “Thousand Brains Theory”?
- Should we represent latent space as continuous or discrete? Would an unbounded length of discrete tokens be sufficient to represent continuous spaces?