



*“Any sufficiently advanced technology is indistinguishable from magic.”*

# Attention Please!

Patrick Michl <patrick.michl@gmail.com>

*Attention Mechanisms*  
in Neural Networks

2022-06-23





*“Any sufficiently advanced technology is indistinguishable from magic.”*

Arthur C. Clarke in *Profiles of the Future* (1962)

# Attention Please!

Patrick Michl <patrick.michl@gmail.com>

*Attention Mechanisms*  
in Neural Networks

2022-06-23



*"Any sufficiently advanced technology is indistinguishable from magic."*

Arthur C. Clarke in *Profiles of the Future* (1962)

```
1 def max_sum_slice(xs):  
2     max_ending = max_so_far = 0  
3  
4  
5  
6
```

# Attention Please!

Patrick Michl <patrick.michl@gmail.com>

*Attention Mechanisms*  
in Neural Networks

2022-06-23





*"Any sufficiently advanced technology is indistinguishable from magic."*

Arthur C. Clarke in *Profiles of the Future* (1962)

```
1 def max_sum_slice(xs):
2     max_ending = max_so_far = 0
3     for x in xs:
4         max_ending = max(0, max_ending + x)
5         max_so_far = max(max_so_far, max_ending)
6     return max_so_far
```

Copilot from OpenAI / GitHub (2022)

# Attention Please!

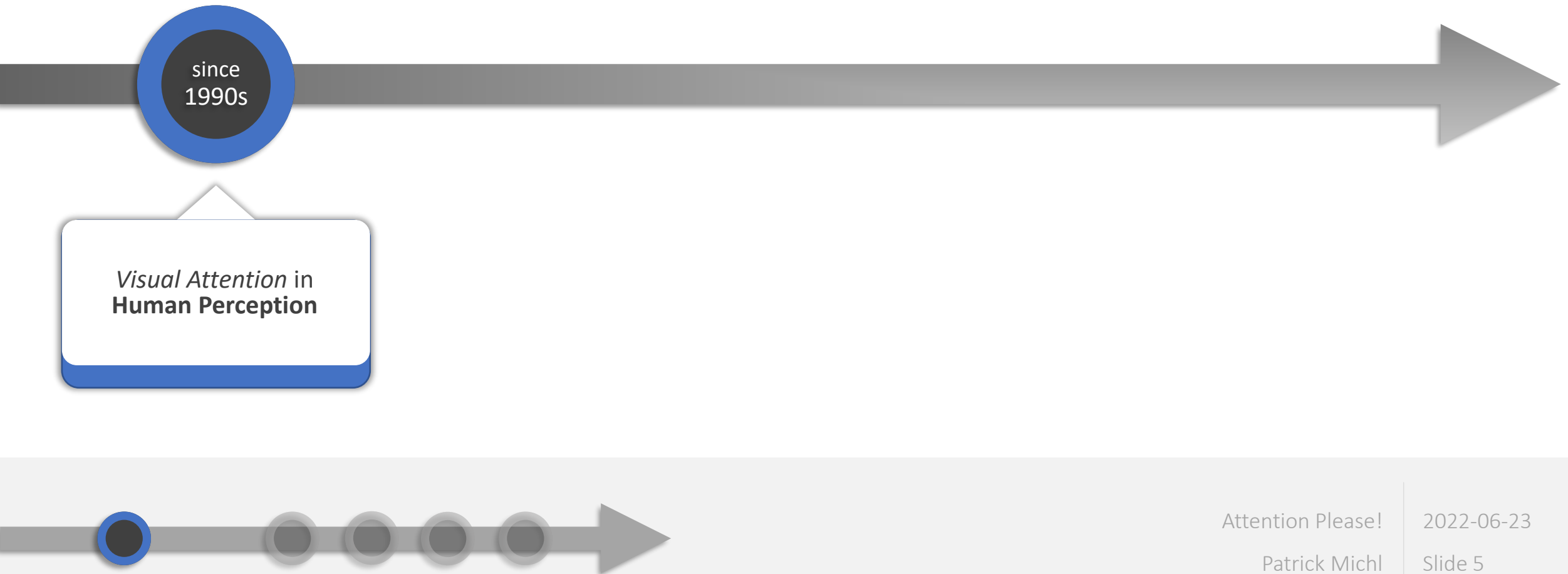
Patrick Michl <patrick.michl@gmail.com>

*Attention Mechanisms*  
in Neural Networks

2022-06-23



# Milestones in Attention Research



# Visual Attention in Human Perception

Visual attention in human perception is based on **selection** and **recognition**

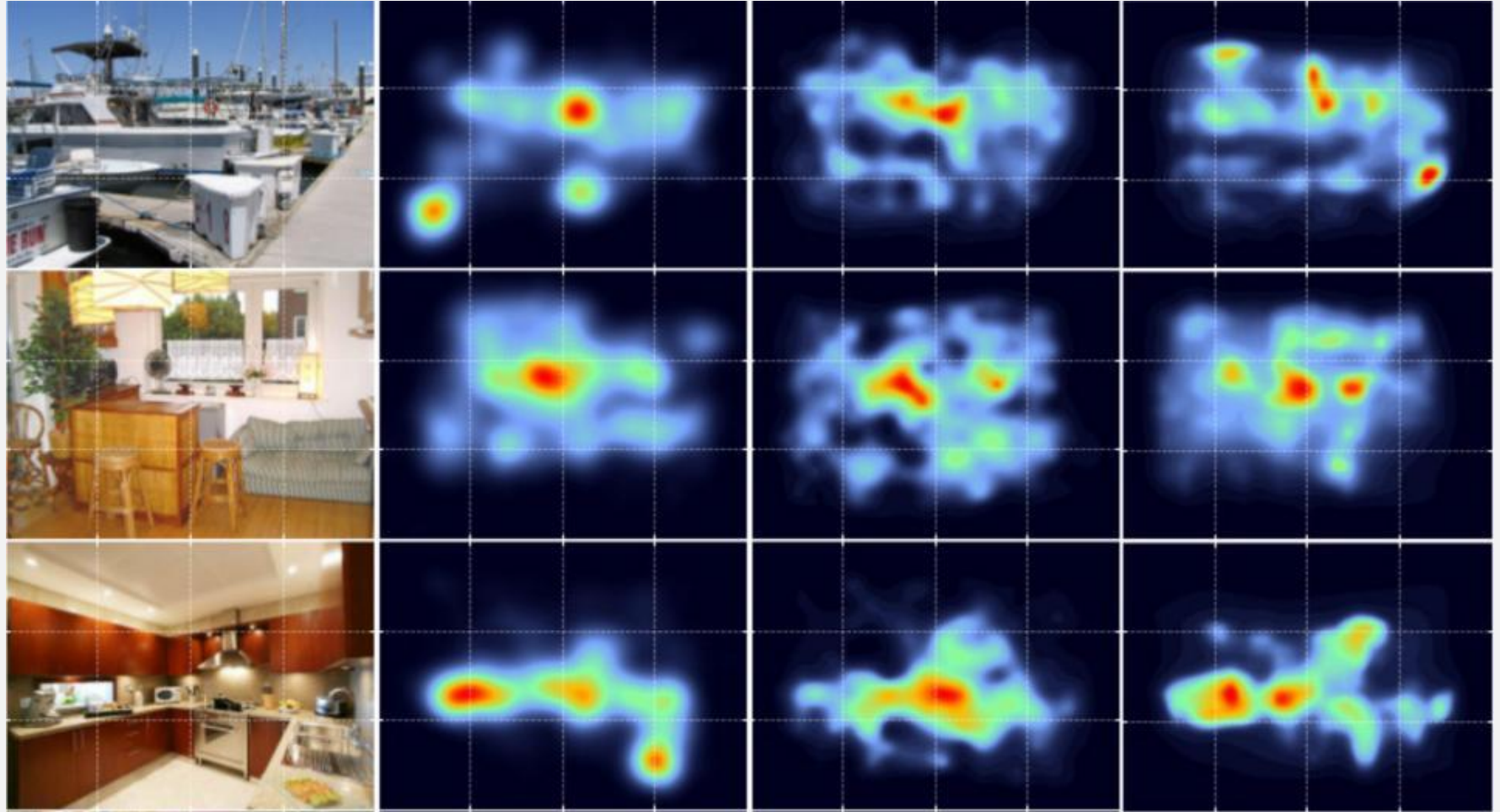


Image source: John Henderson and Taylor Hayes, UC Davis

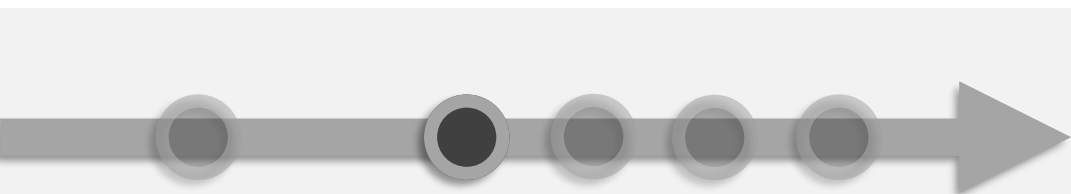
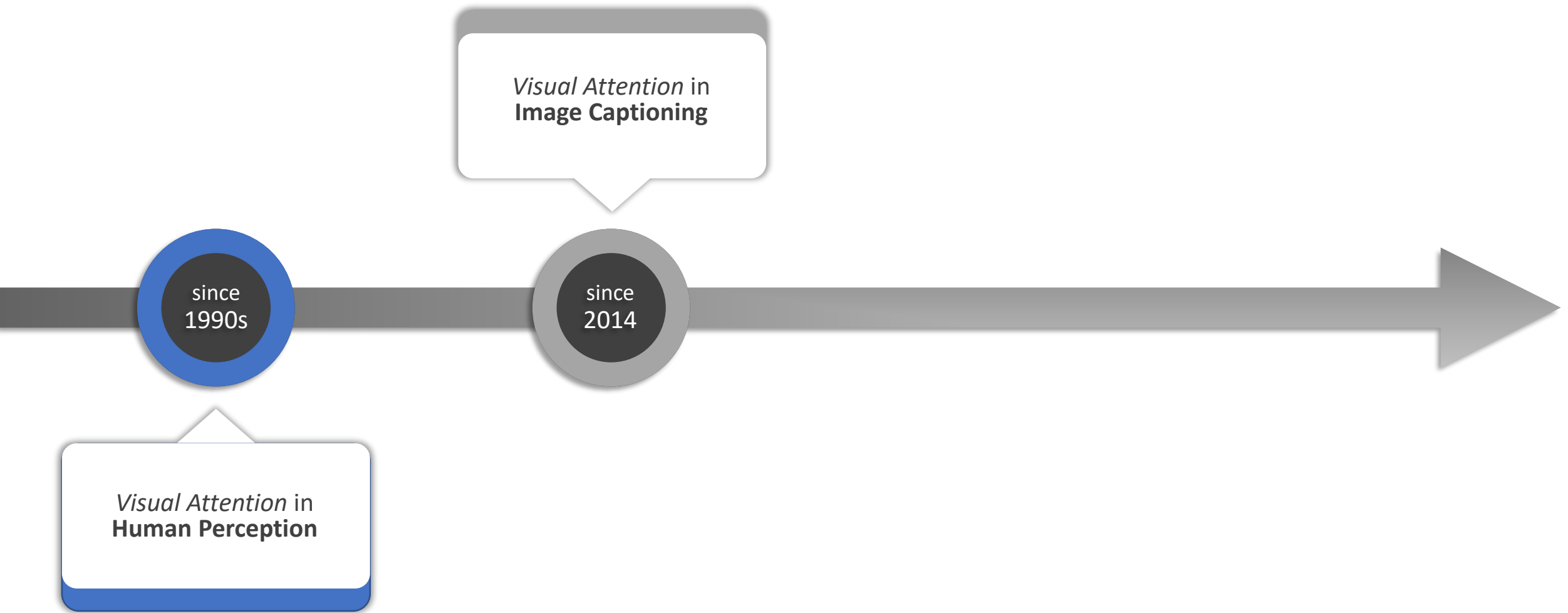


Attention Please!

Patrick Michl

2022-06-23

Slide 6



# Visual Attention in Image Captioning



Spatial

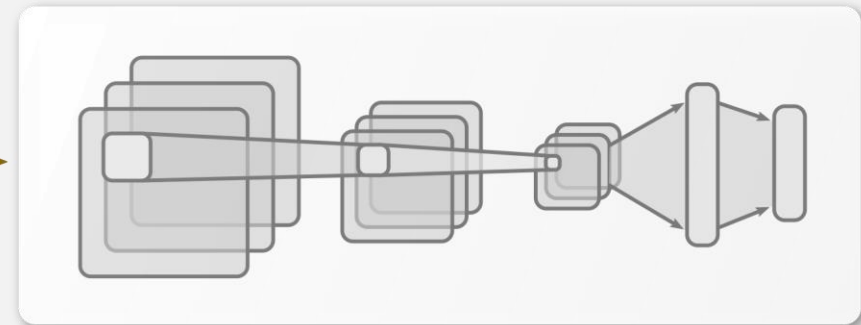
Raw Data

Convolutional Features

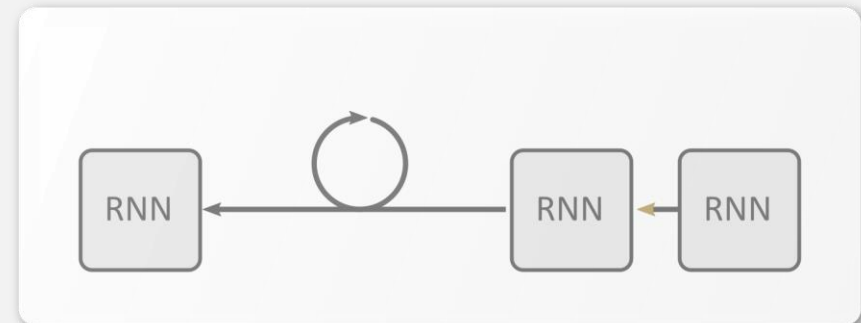
Semantic



ENCODER



DECODER



Attention Please!

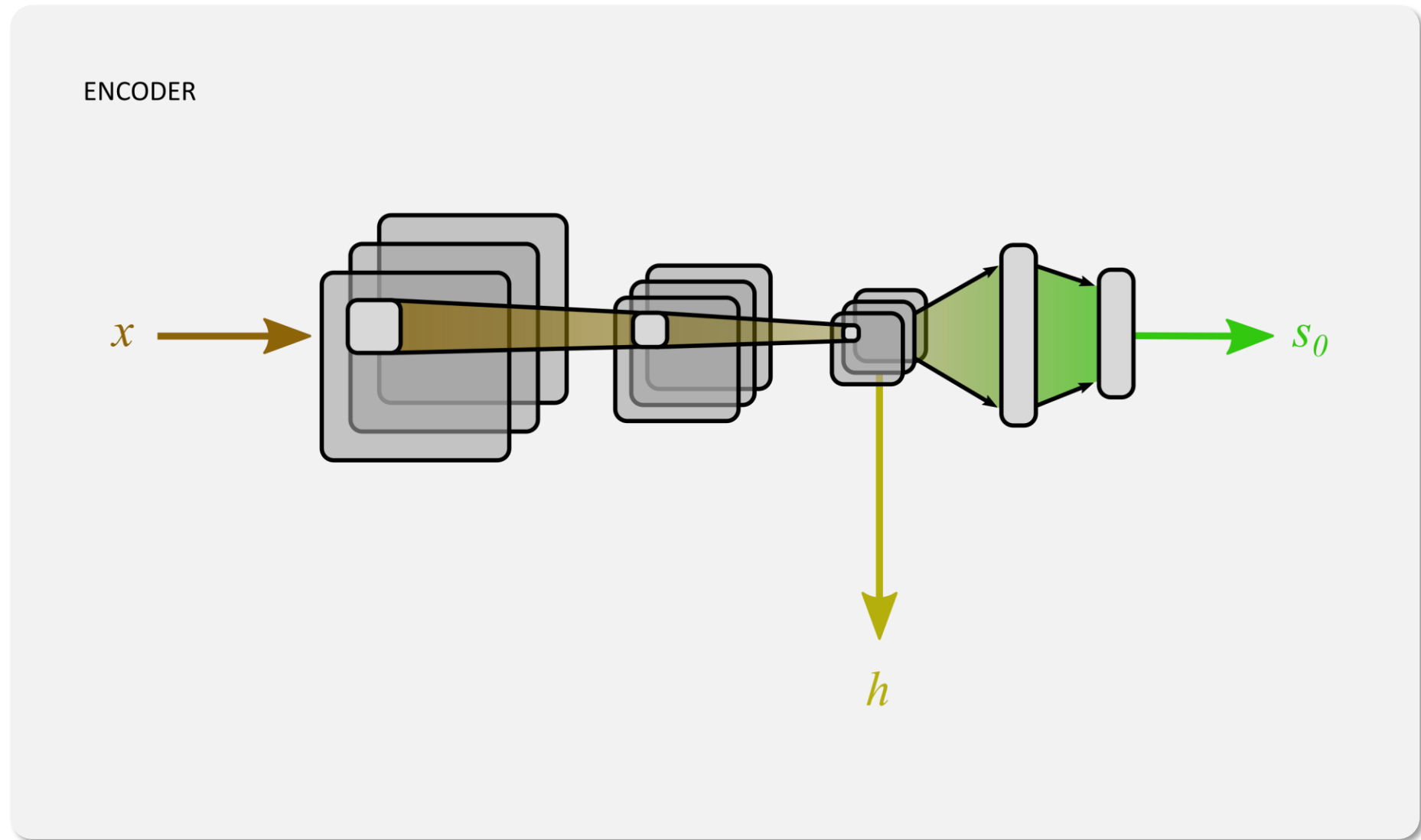
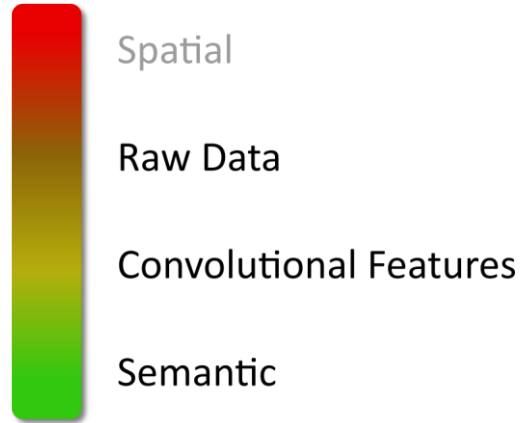
Patrick Michl

2022-06-23

Slide 8



# Visual Attention in Image Captioning



# Visual Attention in Image Captioning



Spatial

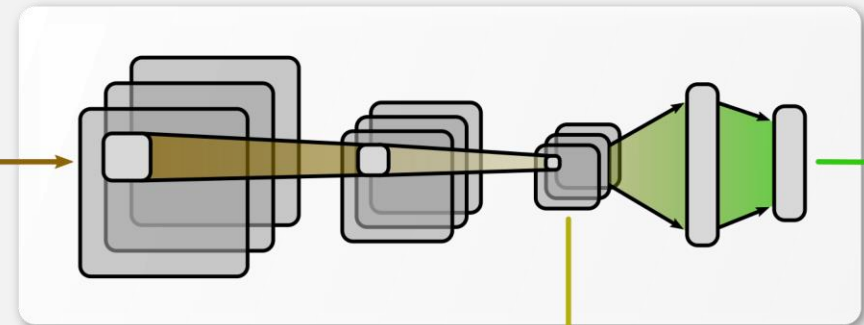
Raw Data

Convolutional Features

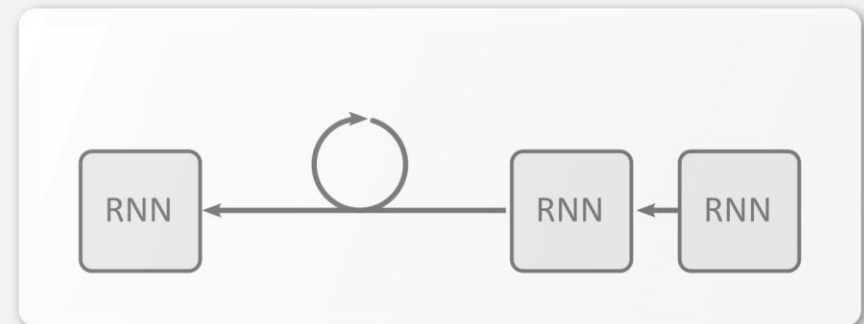
Semantic



ENCODER



DECODER



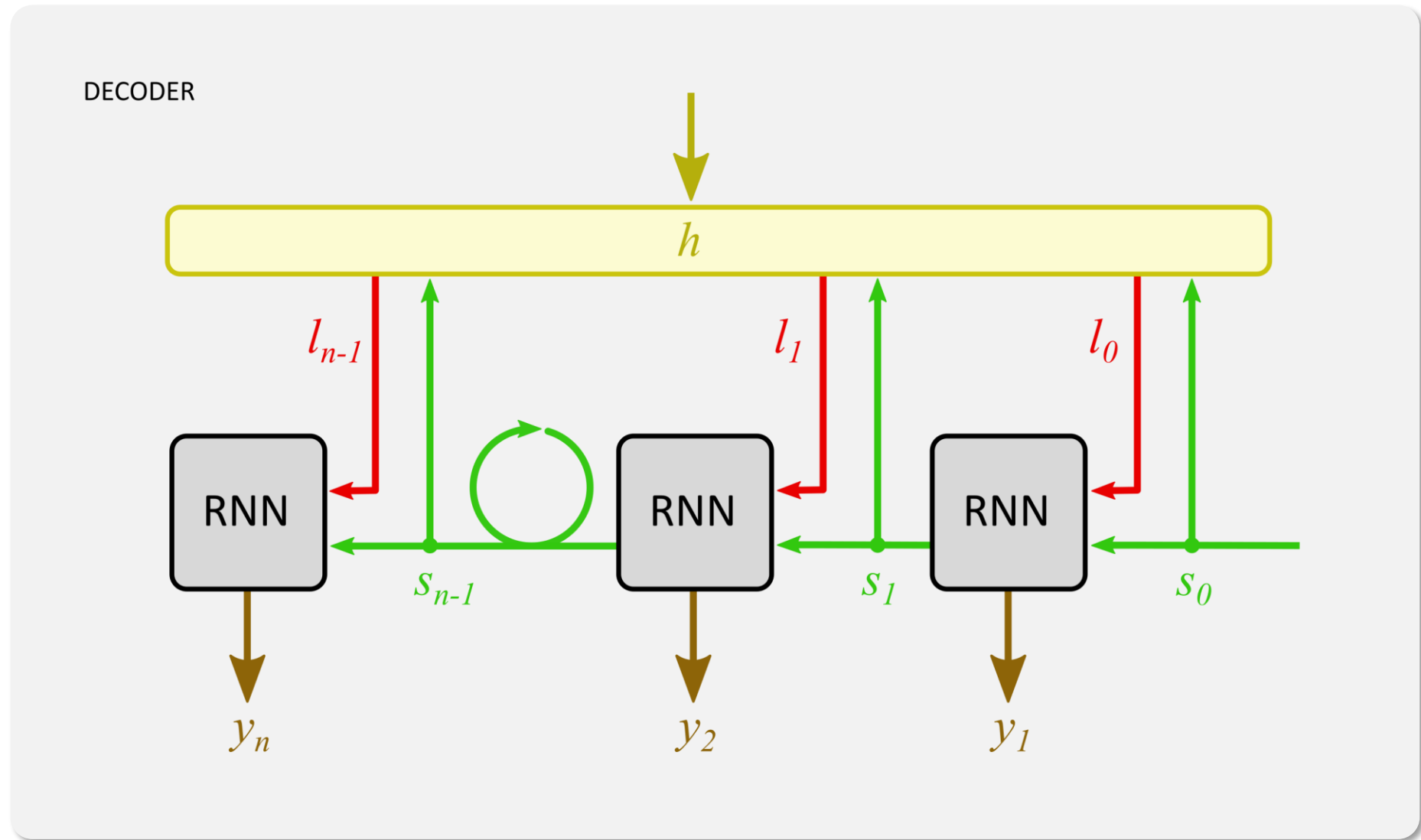
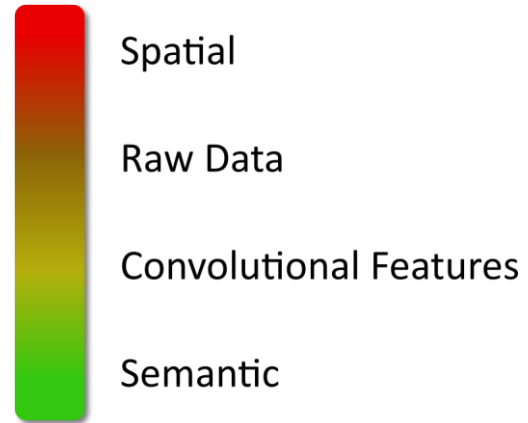
Attention Please!

Patrick Michl

2022-06-23

Slide 10

# Visual Attention in Image Captioning



# Visual Attention in Image Captioning

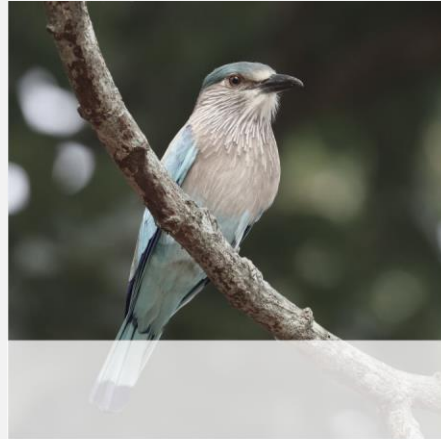


Spatial

Raw Data

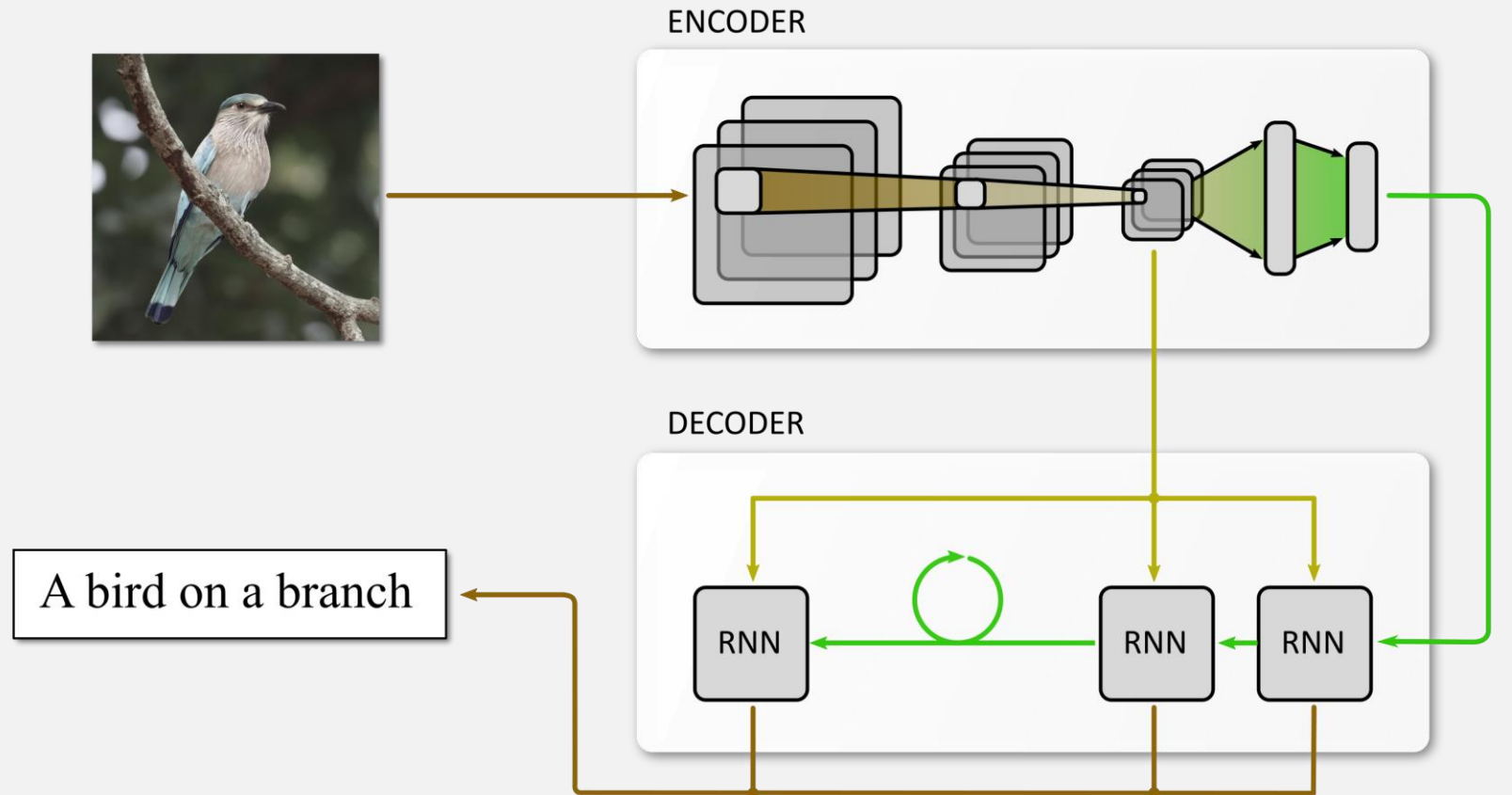
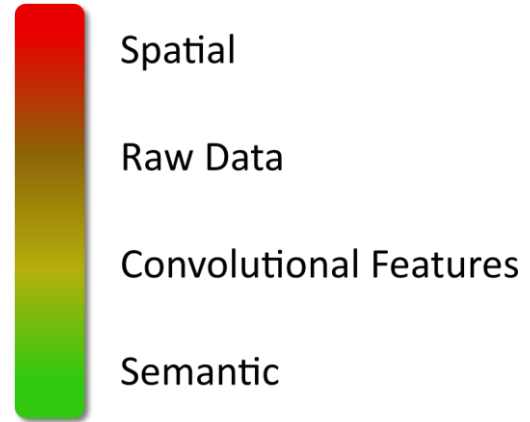
Convolutional Features

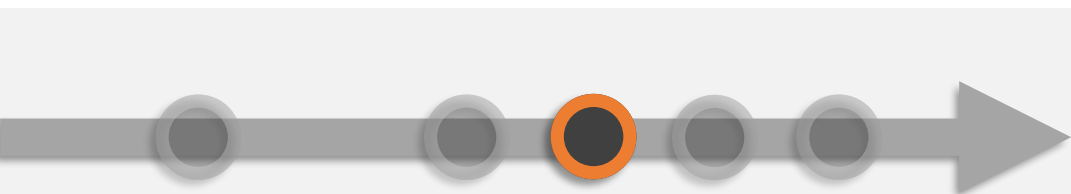
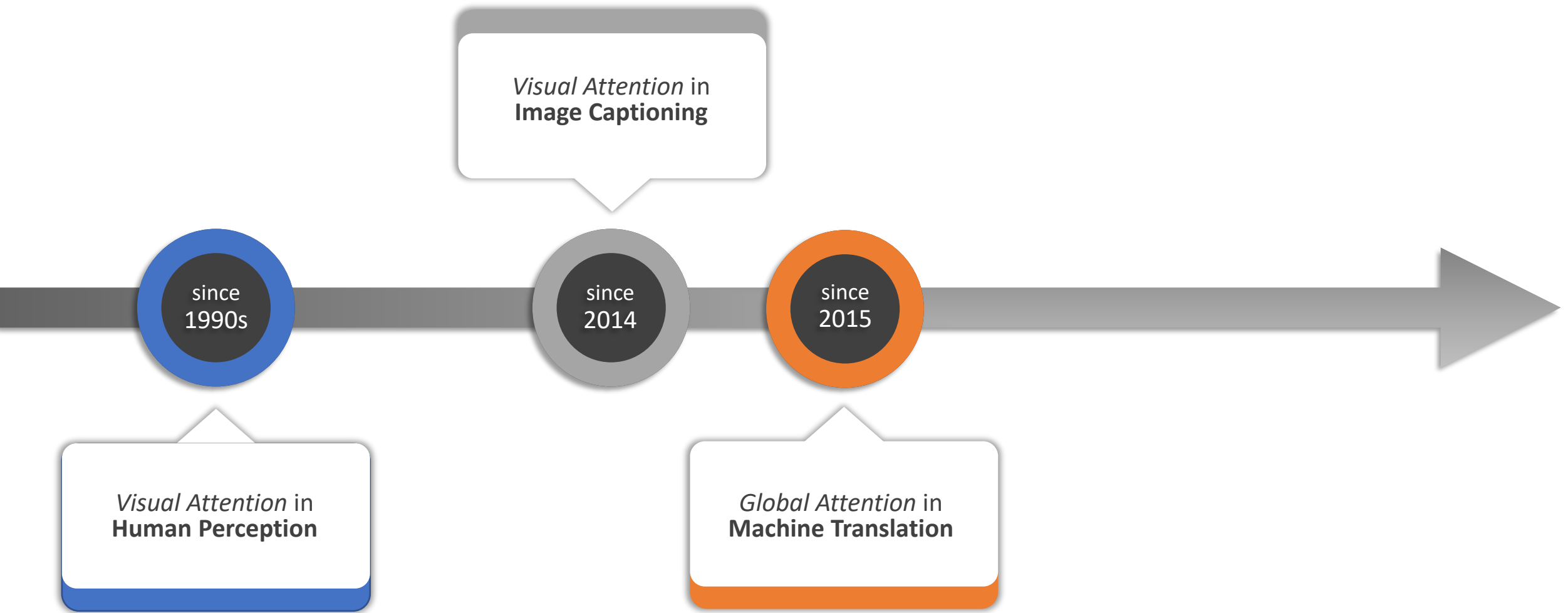
Semantic





# Visual Attention in Image Captioning



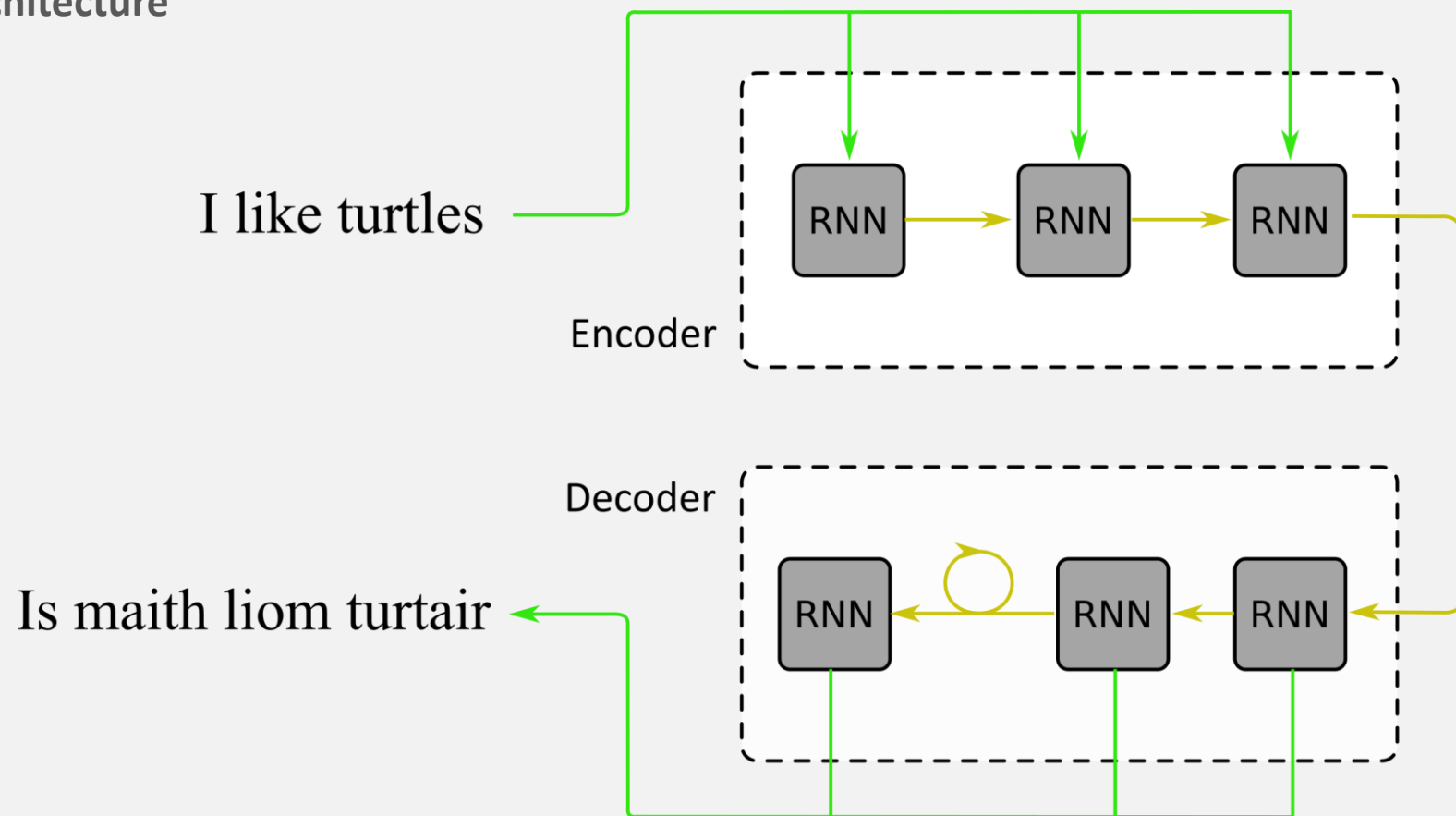


# Global Attention in Machine Translation

Encoder: RNN

Decoder: RNN

## Architecture



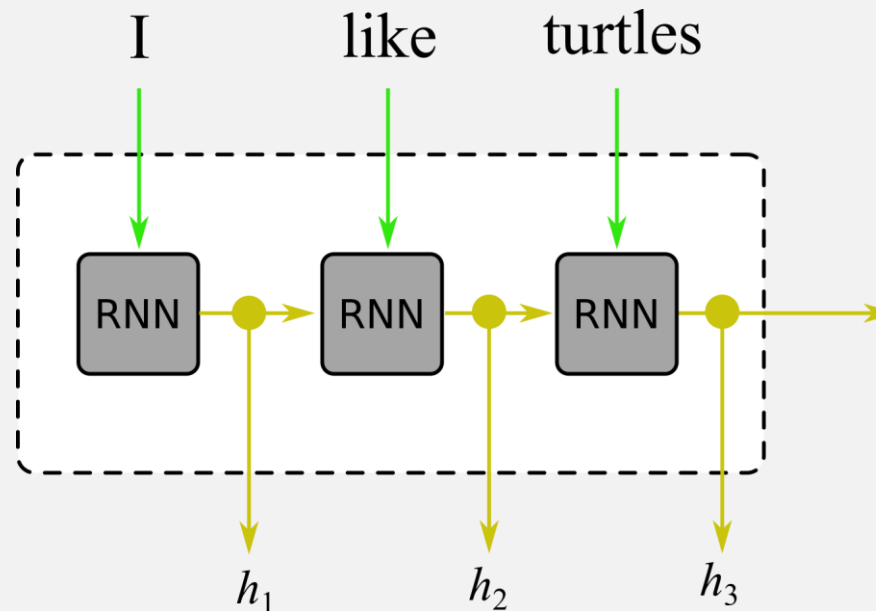
# Global Attention in Machine Translation

**Encoder:** RNN

**Sequence of hidden states**  
encodes the context building  
stack

**Decoder:** RNN

**Encoder**





# Global Attention in Machine Translation

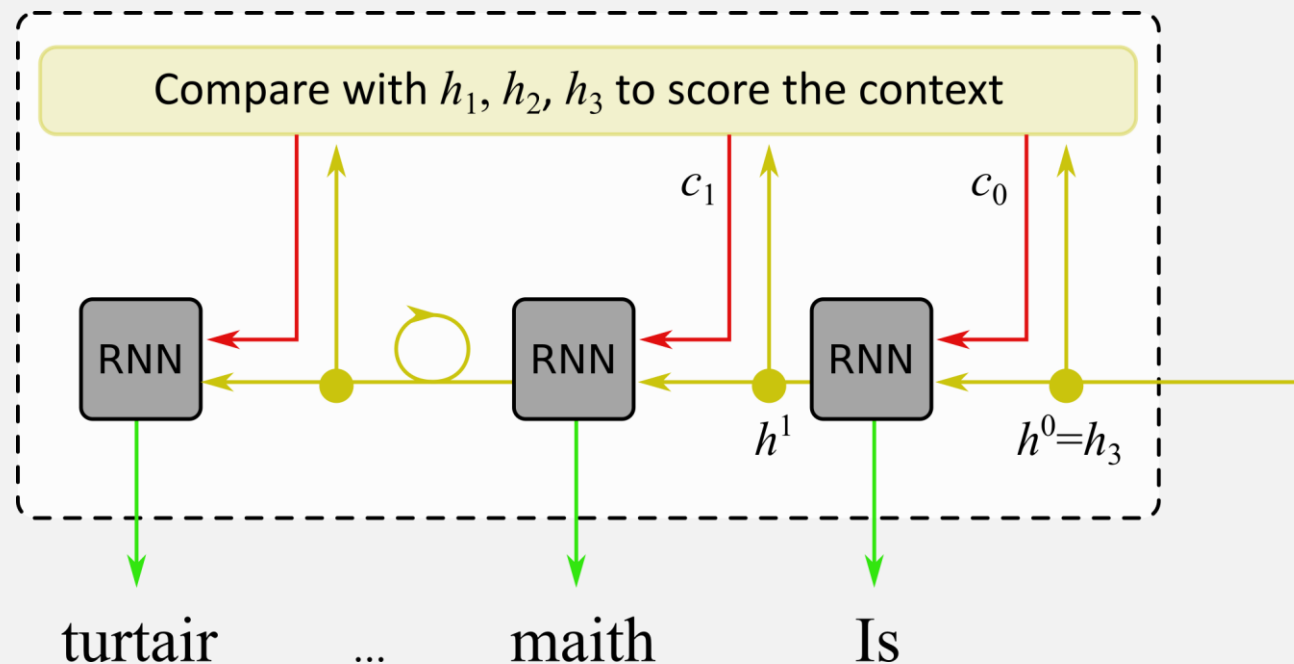
**Encoder: RNN**

**Sequence of hidden states**  
encodes the context building stack

**Decoder: RNN**

**Global Attention**  
allows to recover the context of current outputs

**Decoder**



# Global Attention in Machine Translation

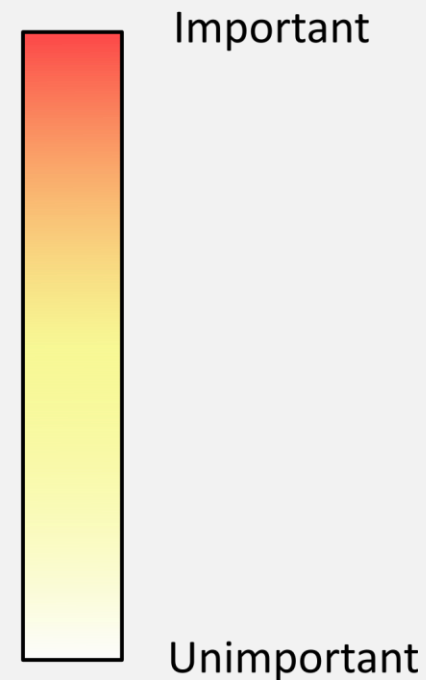
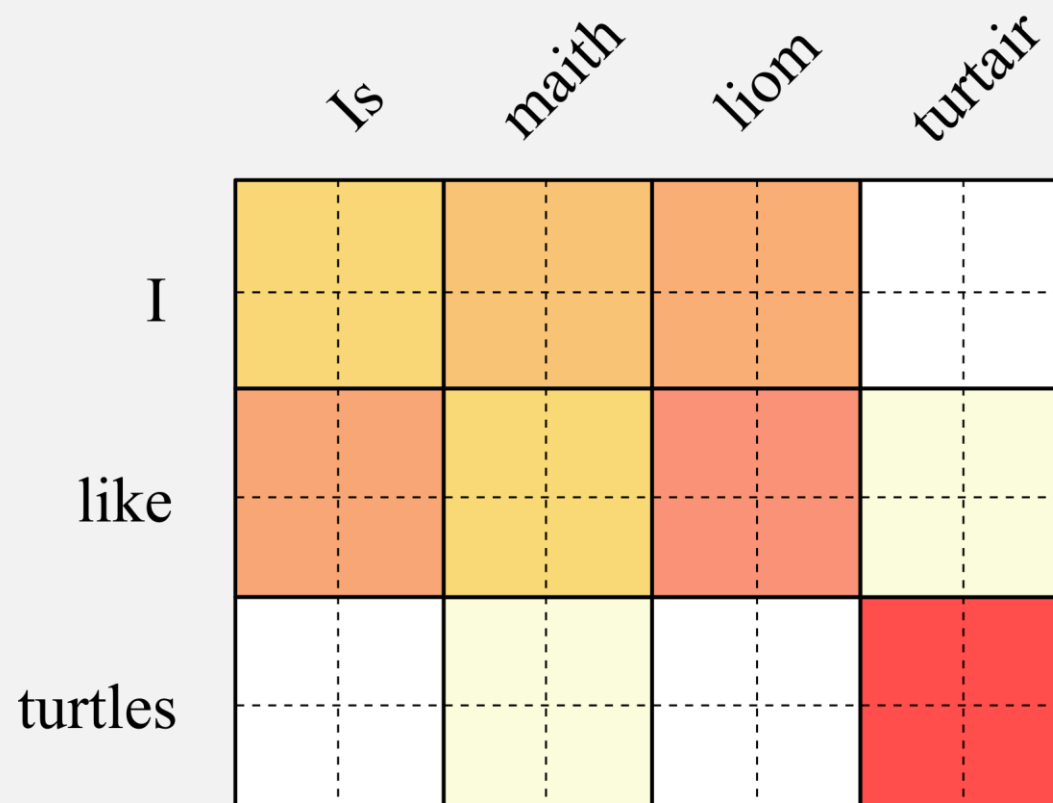
**Encoder: RNN**

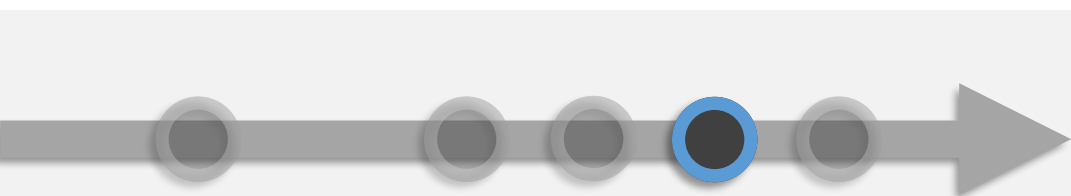
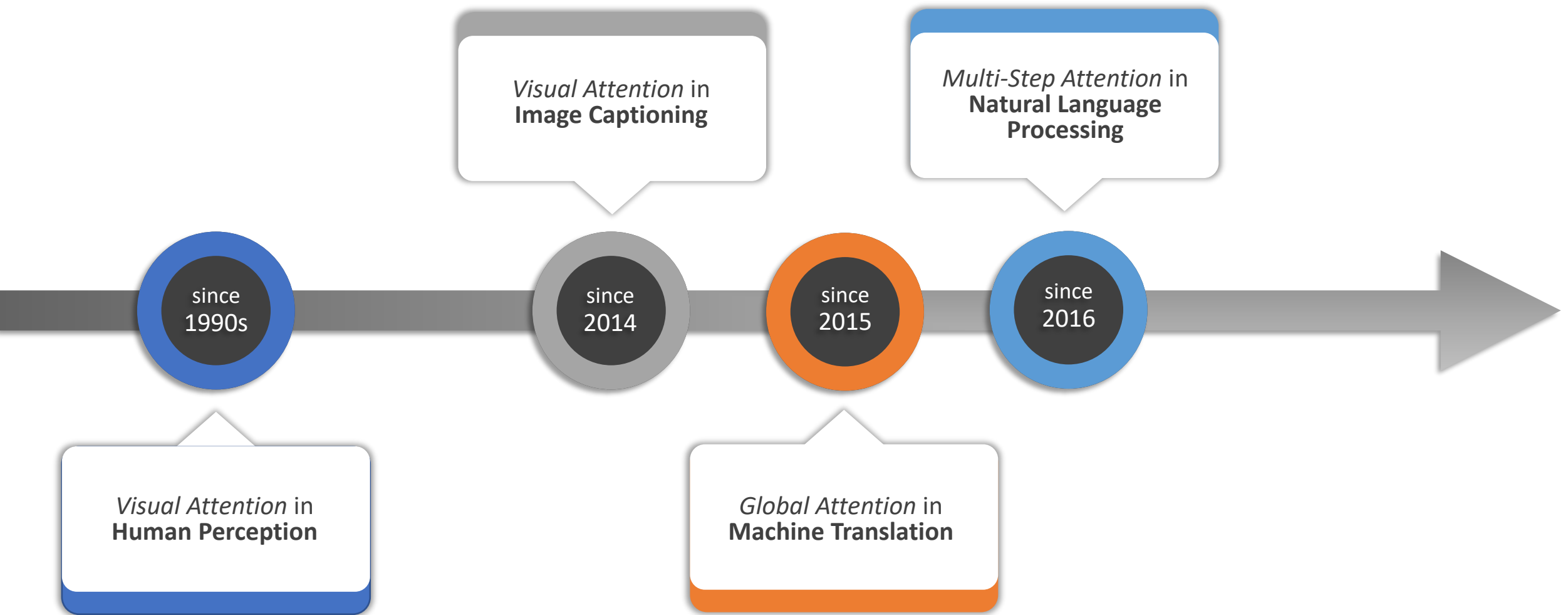
**Sequence of hidden states**  
encodes the context building stack

**Decoder: RNN**

**Global Attention**  
allows to recover the context of current outputs

Context





# Multi-Step Attention in Natural Language Processing

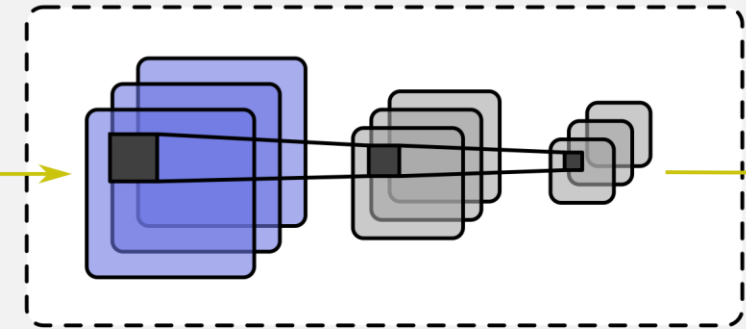
Encoder: CNN

Decoder: CNN

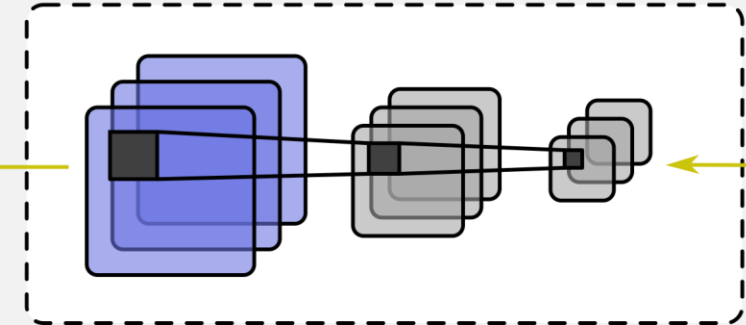
## Architecture

I like turtles,  
so what do I like?

Encoder



Decoder



You like turtles



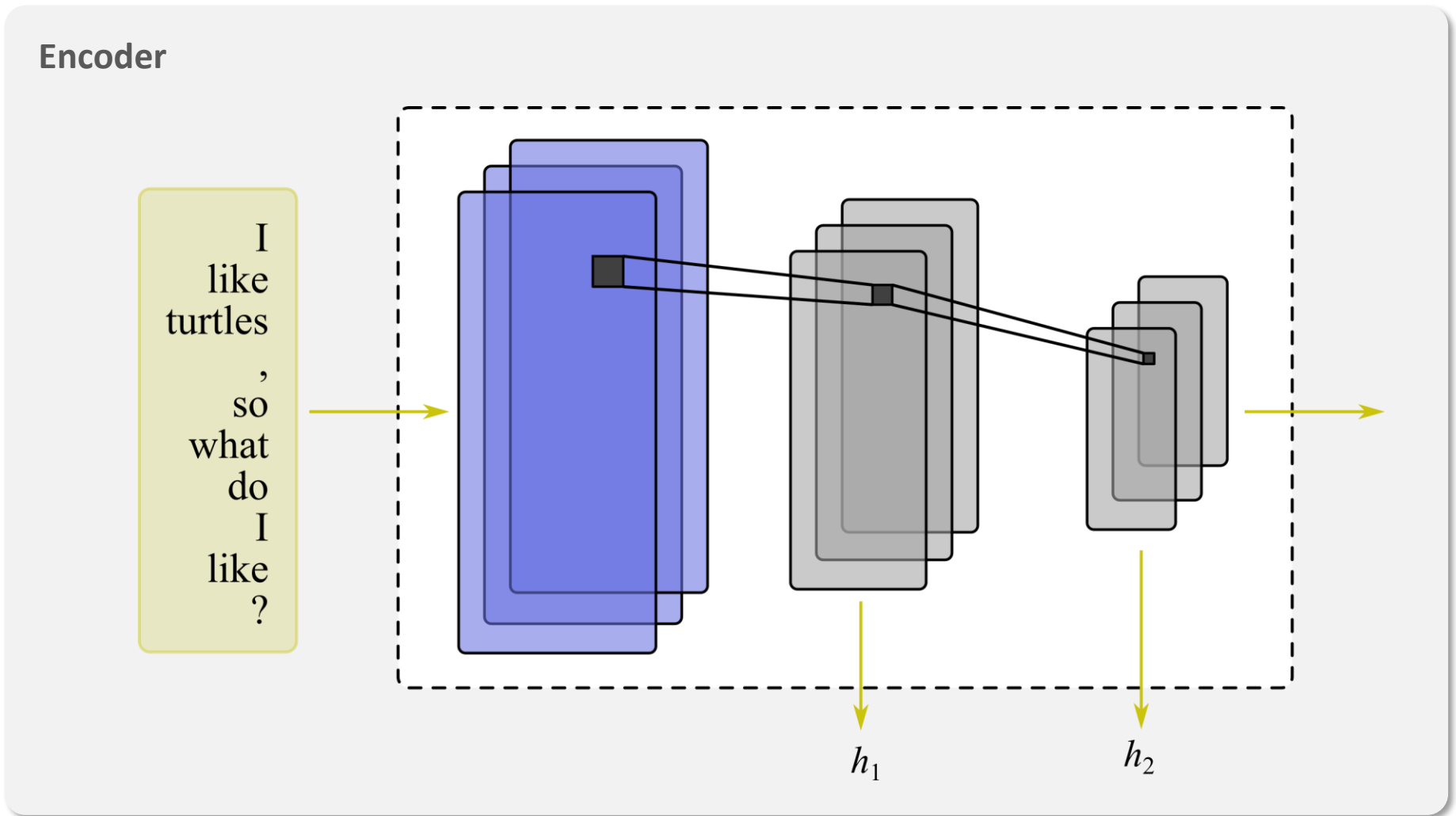


# Multi-Step Attention in Natural Language Processing

**Encoder: CNN**

**Sequence of convolutional  
features** encodes the context  
building stack

**Decoder: CNN**



# Multi-Step Attention in Natural Language Processing

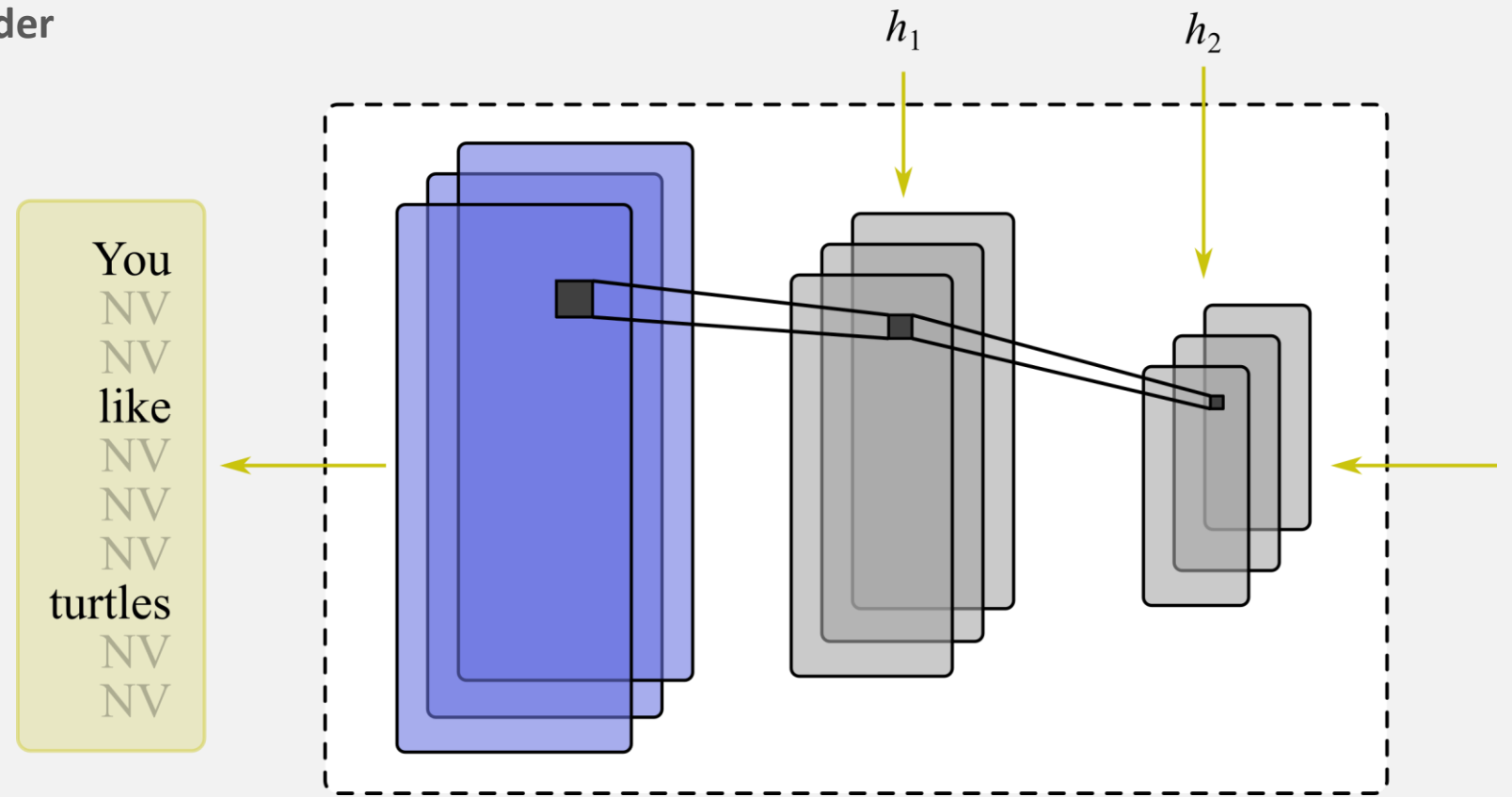
**Encoder: CNN**

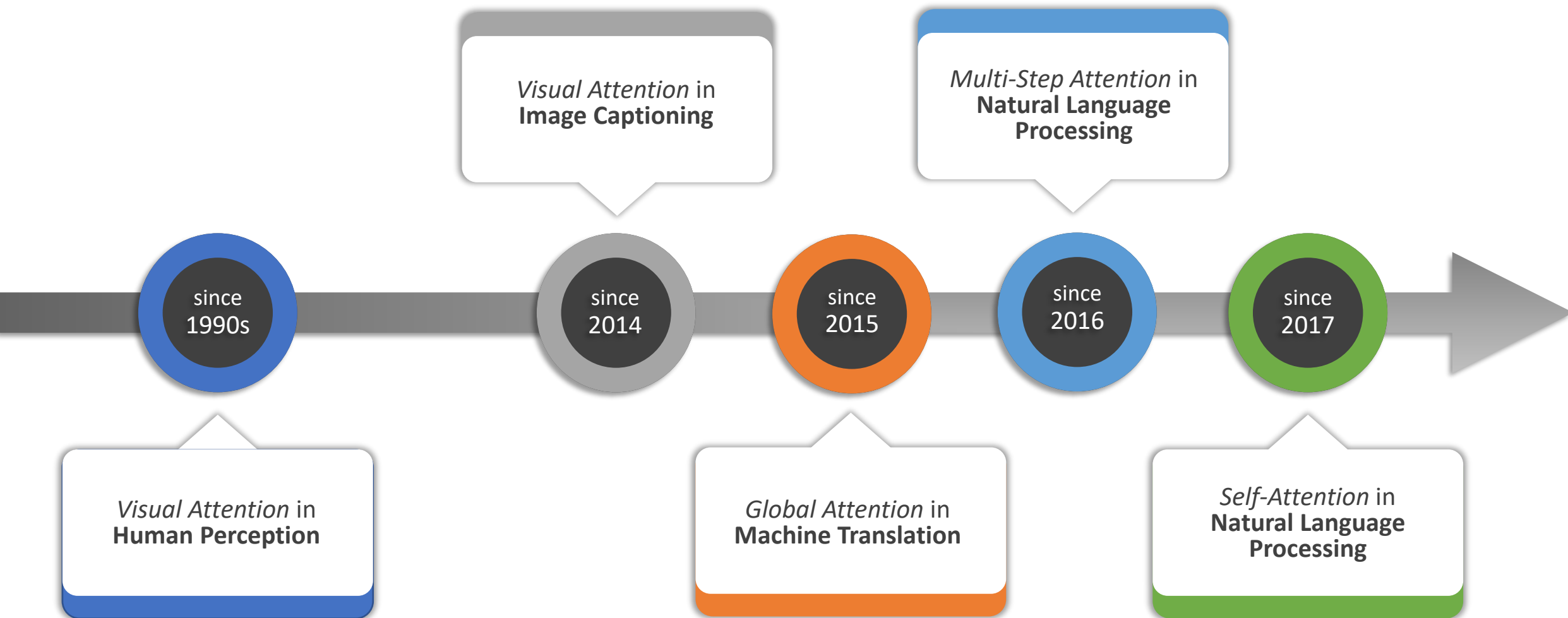
**Sequence of convolutional  
features** encodes the context  
building stack

**Decoder: CNN**

**Multi-Step Attention**  
allows to recover the current  
context of current  
deconvolutional steps

**Decoder**





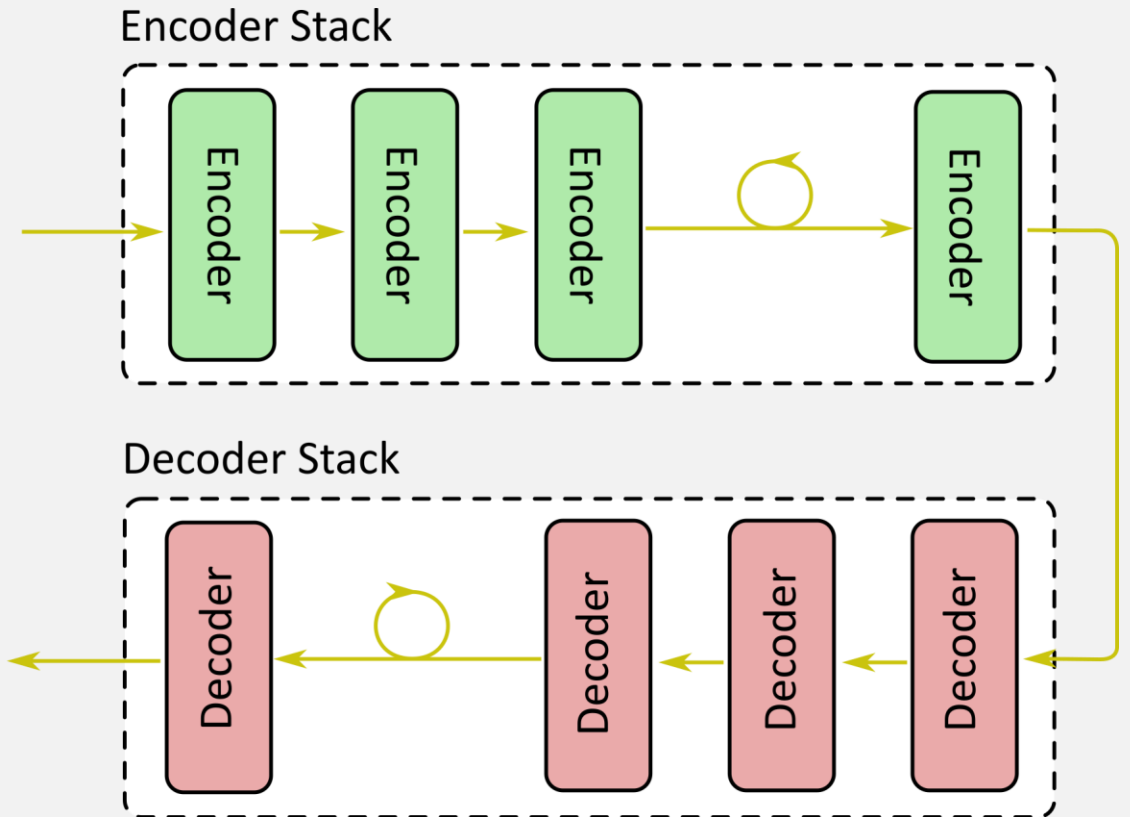
# Self-Attention in Natural Language Processing

**Encoder:** Stacked ANN

**Decoder:** Stacked ANN

## Architecture

Cool kids like turtles  
and he's a cool kid,  
so what does he like?



He likes turtles





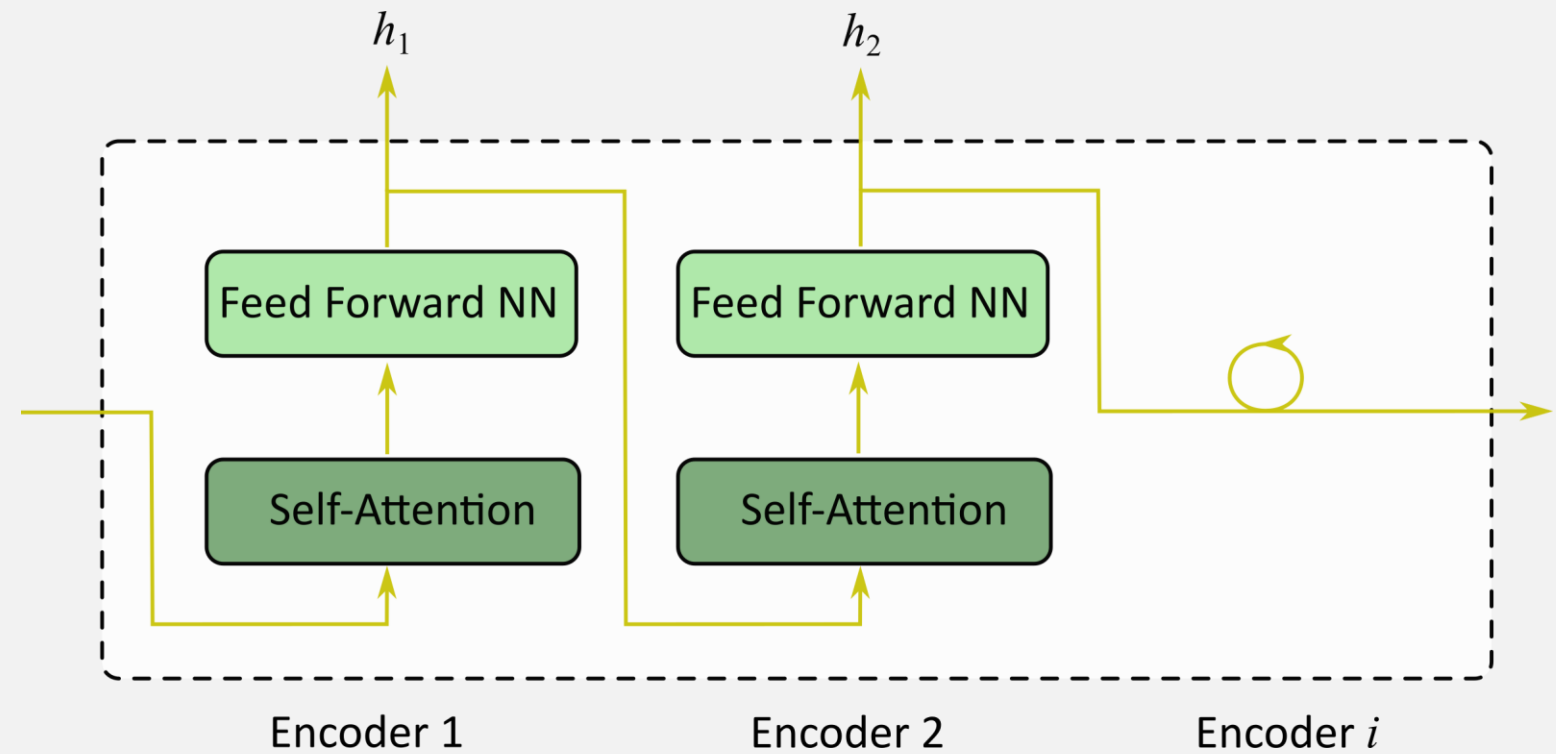
# Self-Attention in Natural Language Processing

**Encoder:** Stacked ANN

Sequence of hidden features captures the encoder context building stack

**Decoder:** Stacked ANN

## Encoder Stack



# Self-Attention in Natural Language Processing

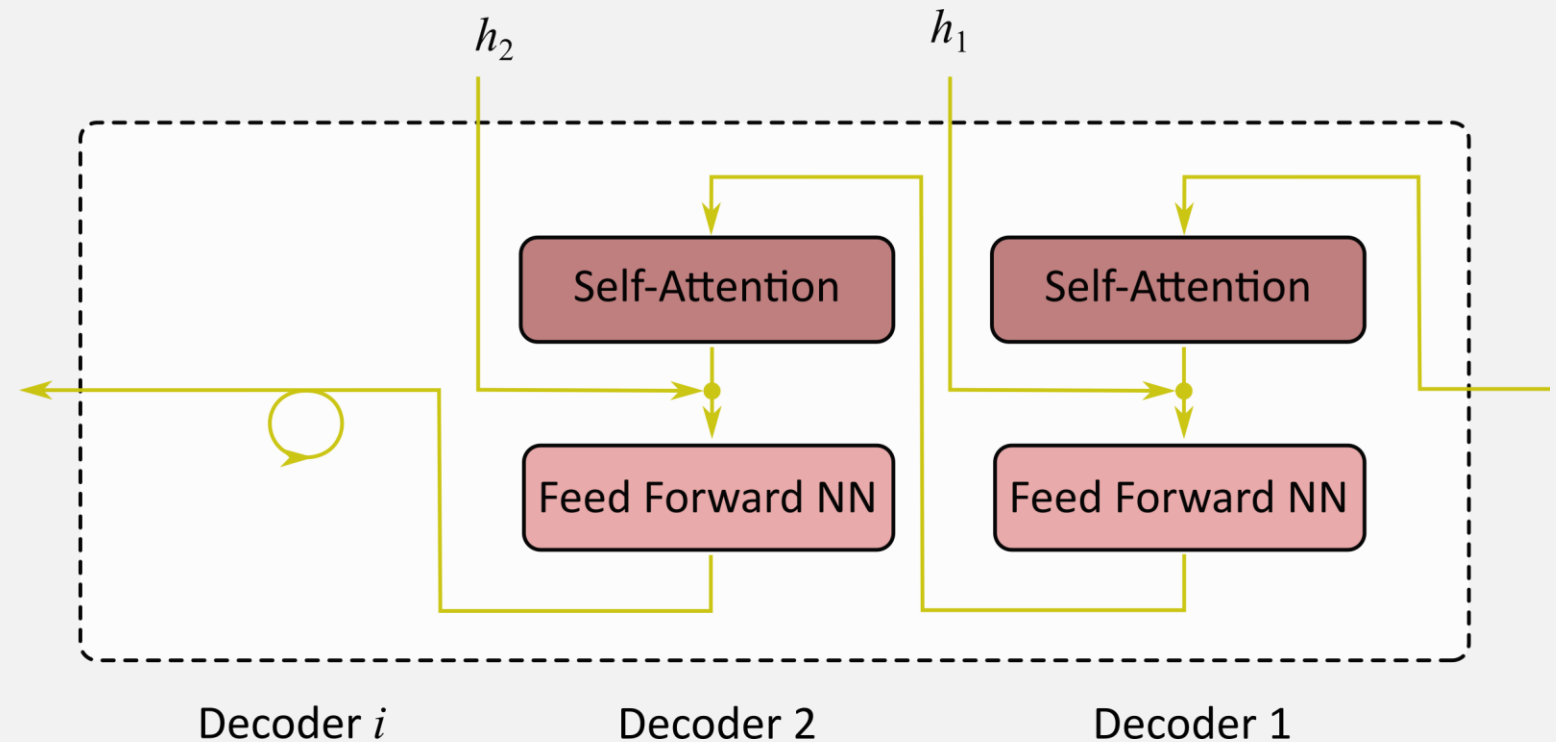
**Encoder:** Stacked ANN

**Sequence of hidden features**  
captures the encoder context  
building stack

**Decoder:** Stacked ANN

**Multi-Step Attention** unrolls  
the context stack to the  
decoder

## Decoder Stack



# Self-Attention in Natural Language Processing

**Encoder:** Stacked ANN

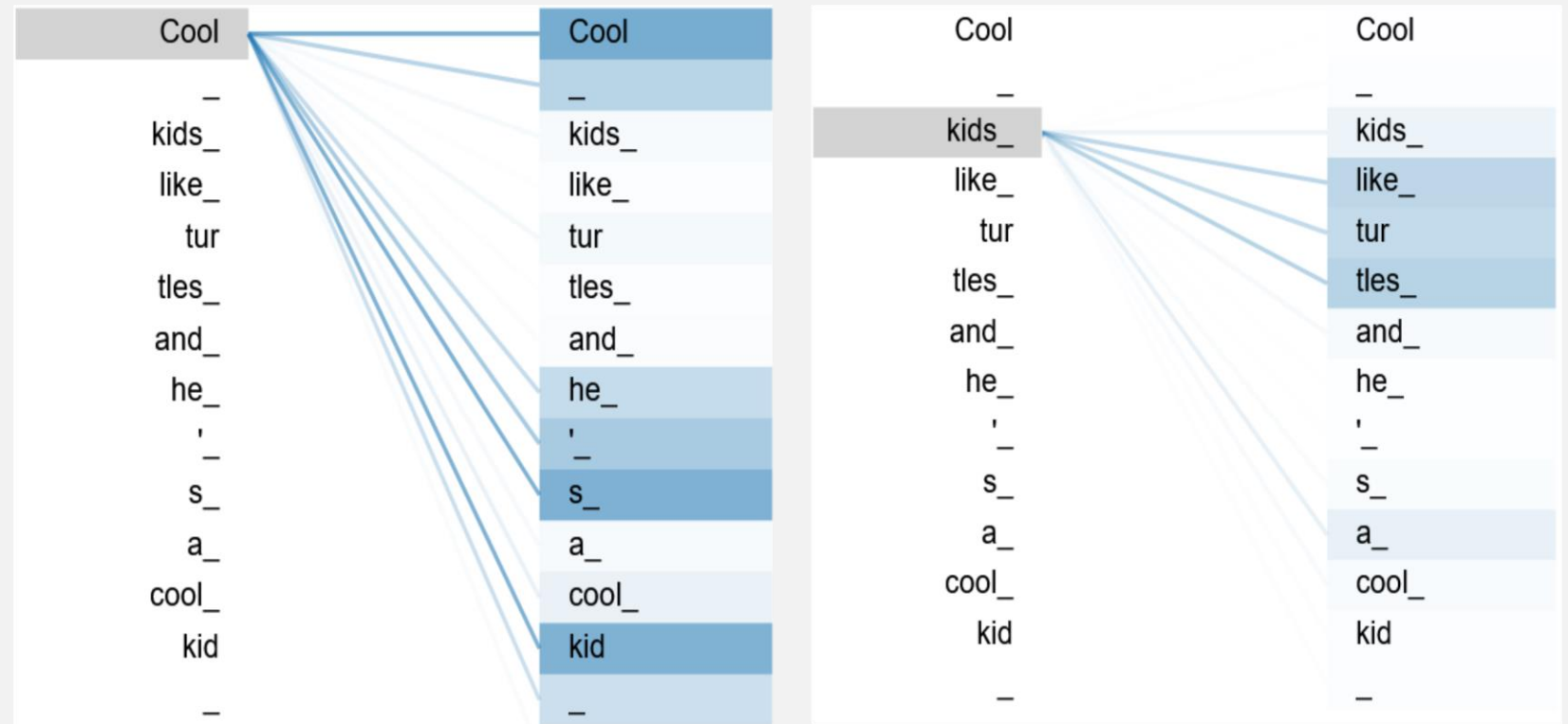
**Sequence of hidden features** captures the encoder context building stack

**Decoder:** Stacked ANN

**Multi-Step Attention** unrolls the context stack to the decoder

-----  
**Self-attention** aggregates context dependencies within the inputs

## Self-Attention



# Summary

## #1

Human Perception is based on the dynamics between ***selection*** and ***recognition***

## #2

Attention mechanisms immitate this behaviour by using intermediate states, that entangle *context* with *semantic* information

## #3

The incorporation of context provides *dynamic features* that are context specific and therefore improve the model performance

## #4

After all - attention mechanisms can also help to *understand the decisions* of deep networks

# Summary

## #1

Human Perception is based on the dynamics between *selection* and *recognition*

## #2

Attention mechanisms immitate this behaviour by using intermediate states, that entangle *context* with *semantic* information

## #3

The incorporation of context provides *dynamic features* that are context specific and therefore improve the model performance

## #4

After all - attention mechanisms can also help to *understand the decisions* of deep networks

# Summary

## #1

Human Perception is based on the dynamics between *selection* and *recognition*

## #2

Attention mechanisms immitate this behaviour by using intermediate states, that entangle *context* with *semantic* information

## #3

The incorporation of context provides **dynamic features** that are context specific and therefore improve the model performance

## #4

After all - attention mechanisms can also help to **understand the decisions** of deep networks

# Summary

## #1

Human Perception is based on the dynamics between *selection* and *recognition*

## #2

Attention mechanisms immitate this behaviour by using intermediate states, that entangle *context* with *semantic* information

## #3

The incorporation of context provides **dynamic features** that are context specific and therefore improve the model performance

## #4

After all - attention mechanisms can also help to **understand the decisions** of deep networks



Thank you for your attention!

