

Shannon Revisited: N-grams and autoregressive



Prediction and Entropy of Printed English

By C. E. SHANNON

(Manuscript Received Sept. 15, 1950)

A new method of estimating the entropy and redundancy of a language is described. This method exploits the knowledge of the language statistics possessed by those who speak the language, and depends on experimental results in prediction of the next letter when the preceding text is known. Results of experiments in prediction are given, and some properties of an ideal predictor are developed.

Markov Chain

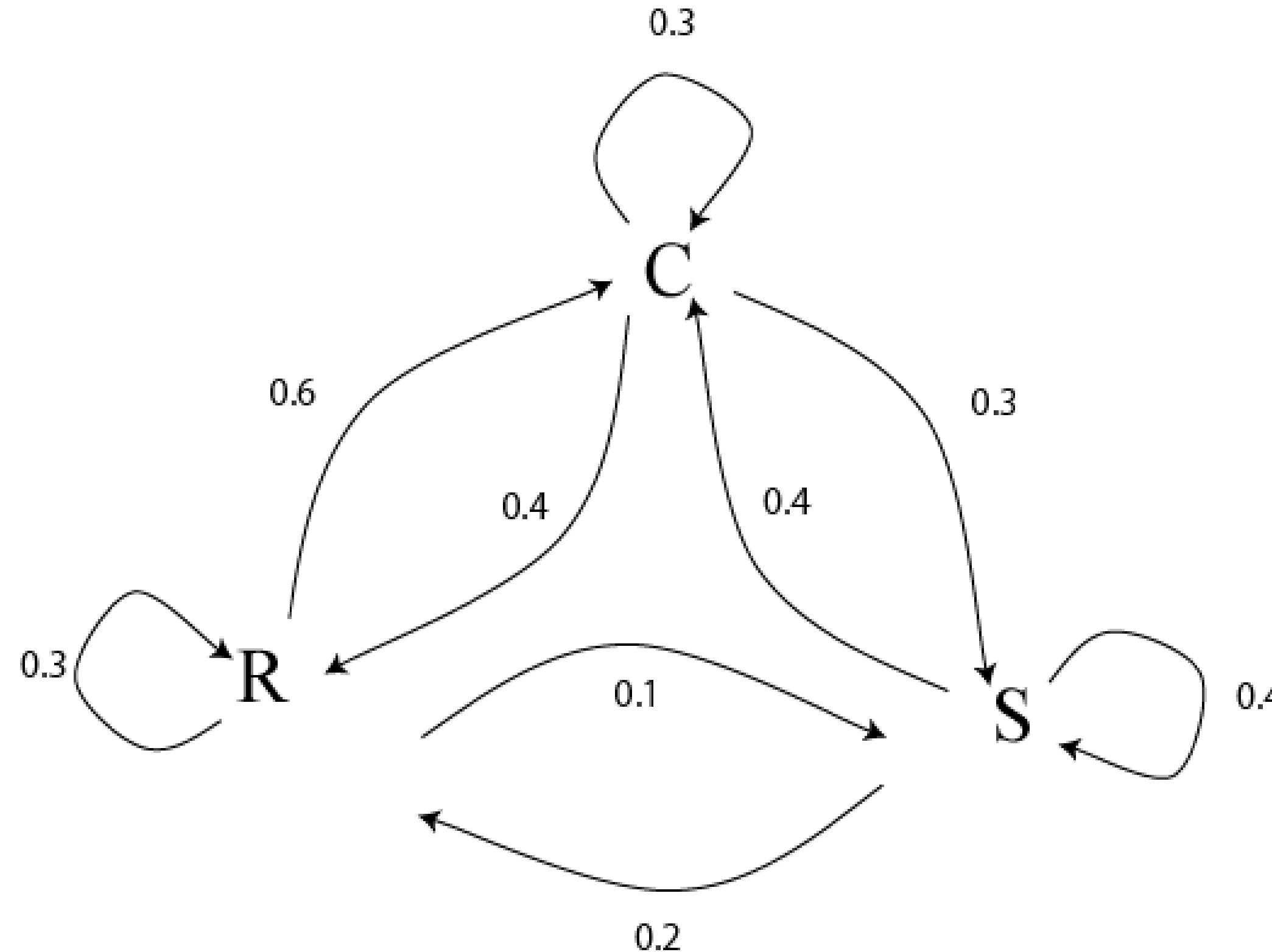
Let's predict weather:

- Given today's weather only, we want to know tomorrow's
- Suppose weather can only be {Sunny, Cloudy, Raining}

The “Weather Channel” algorithm:

- Over a long period of time, record:
 - How often S followed by R
 - How often S followed by S
 - Etc.
- Compute percentages for each state:
 - $P(R|S)$, $P(S|S)$, etc.
- Predict the state with highest probability!
- It's a Markov Chain

Markov Chain



$$\begin{pmatrix} 0.3 & 0.6 & 0.1 \\ 0.4 & 0.3 & 0.3 \\ 0.2 & 0.4 & 0.4 \end{pmatrix}$$

What if we know today and yesterday's weather?

Text Synthesis

[Shannon,'48] proposed a way to generate English-looking text using N-grams:

- Assume a generalized Markov model
- Use a large text to compute prob. distributions of each letter given N-1 previous letters
- Starting from a seed repeatedly sample this Markov chain to generate new letters
- Also works for whole words

WE NEED TO EAT CAKE

Results (using alt.singles corpus):

- “As I've commented before, really relating to someone involves standing next to impossible.”
- “One morning I shot an elephant in my arms and kissed him.”
- “I spent an interesting evening recently with a grain of salt”

Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was still red

Bob Dylan, *Tangled up in Blue*

slide from Steve Seitz's [video](#)

Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was still red

Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was still red

Early one morning the sun was shining laying in bed
her hair still red

Wondering if she had changed at all if

```
graph TD; sun[the sun] --- was[was]; was --- shining[shining]; shining --- laying[laying in bed]; hair[her hair] --- still[still red];
```

Early one morning the sun was shining

her hair still red

Wondering if she had changed at all if

laying in bed

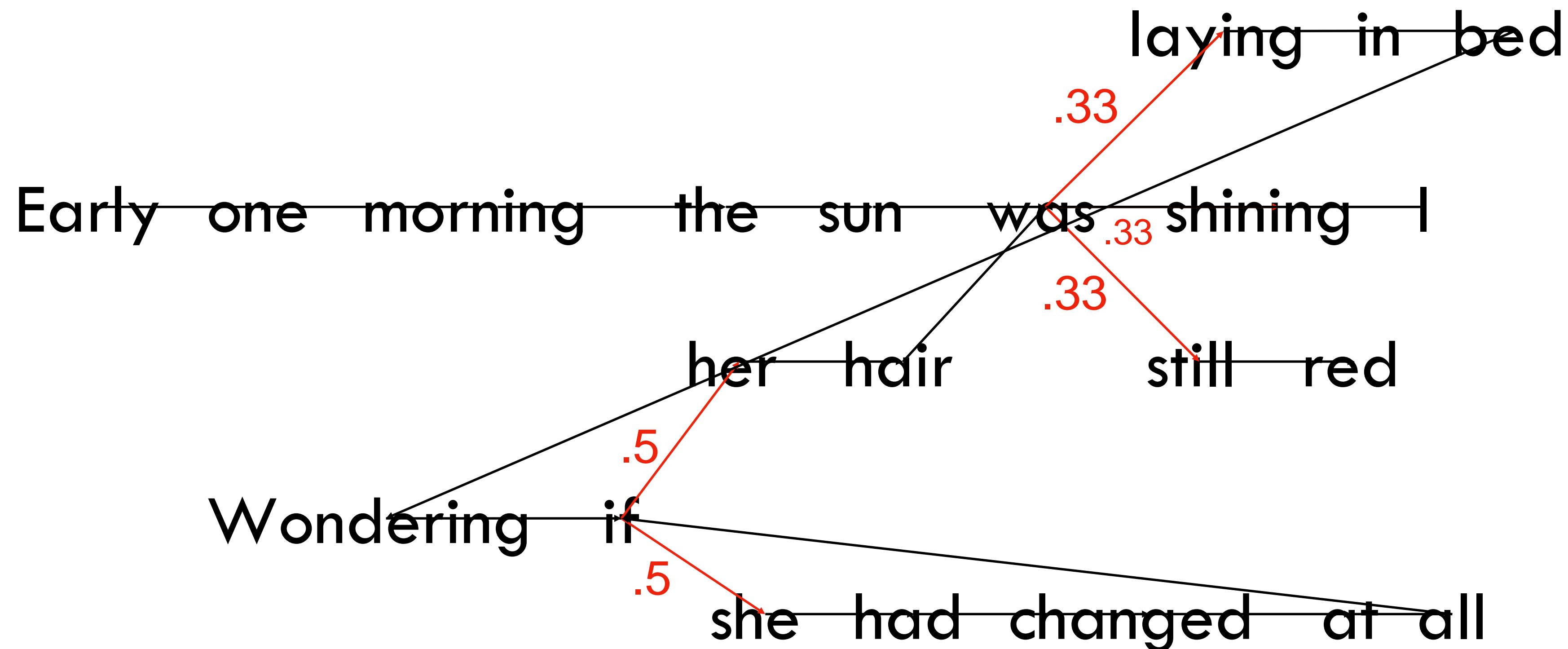
```
graph TD; A[the sun was shining] --- B[her hair]; B --- C[still red]; D[Wondering if she had changed at all if]; D --- E[if]; D --- F[if]; G[laying in bed]
```

Early one morning the sun was shining

Wondering if she had changed at all

her hair still red

laying in bed



Language Model

Early one morning the sun was shining I
laying in bed still red
Wondering if her hair was still red
she had changed at all

Early

Early → one morning the sun was shining I

laying in bed

her hair still red

Wondering if she had changed at all

```
graph TD; Early((Early)) --> one((one)); one --- morning((morning)); morning --- the((the)); the --- sun((sun)); sun --- was((was)); was --- shining((shining)); shining --- I((I)); her((her)) --- hair((hair)); still((still)) --- red((red)); Wondering((Wondering)) --- if((if)); if --- she((she)); she --- had((had)); had --- changed((changed)); changed --- at((at)); at --- all((all))
```

Early one

Early one → morning the sun was shining I

laying in bed

her hair still red

Wondering if she had changed at all

The diagram illustrates a process of word modification or replacement. It starts with the sentence "Early one morning the sun was shining I". A red dot is placed above the word "Early". Several words are crossed out with lines pointing to alternative words: "one" points to "morning"; "was" points to "shining"; "I" points to "bed"; "the sun" points to "her hair"; and "Wondering if" points to "she had changed at all". The word "was" also has a horizontal line extending from its right side.

Early one morning

Early one morning → the sun was shining I

Wondering if her hair still red

she had changed at all

```
graph LR; Dot(( )) --- Early[Early]; Dot --- The[the]; The --- Sun[sun]; Was[was] --- Shining[shining]; Her[her] --- Hair[hair]; She[she] --- Changed[changed]; If;if --- Hair; At[at] --- All[all]
```

Early one morning the

Early one morning the sun was shining I
laying in bed
her hair still red
Wondering if she had changed at all

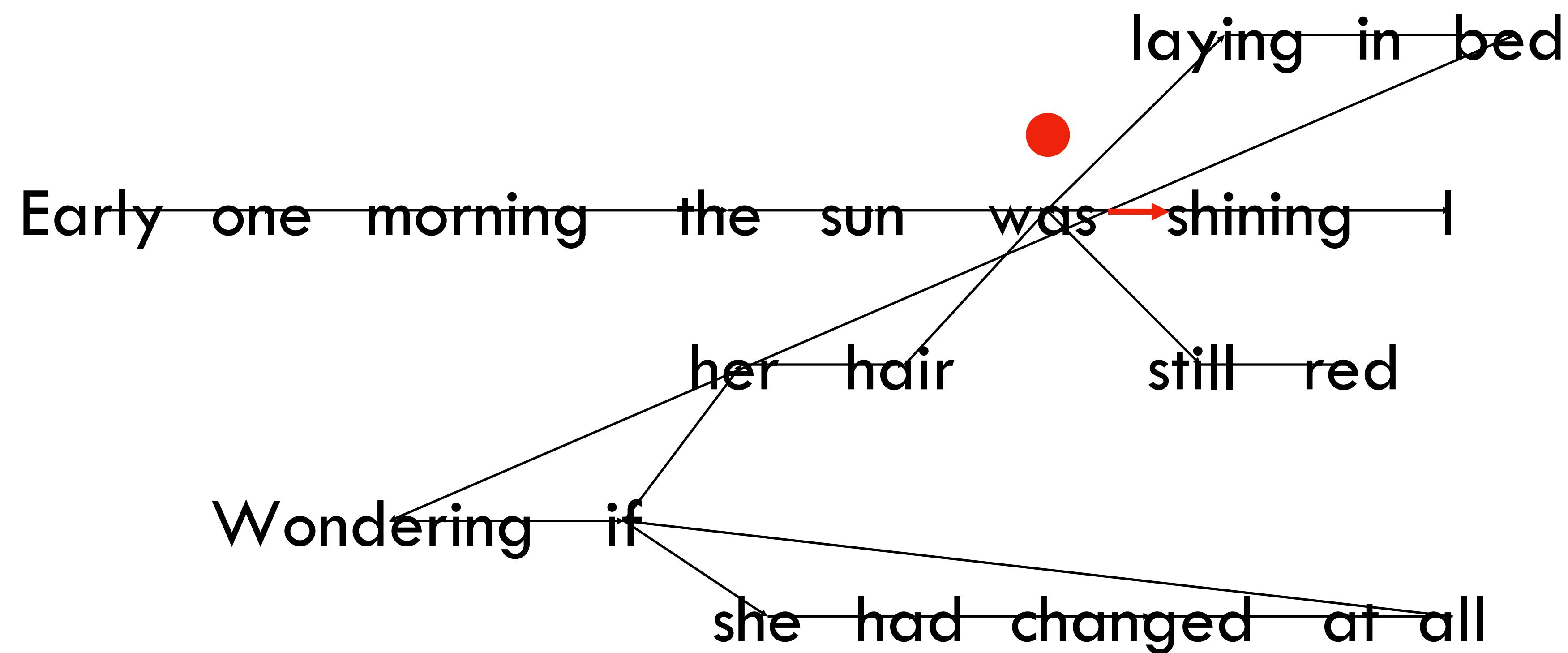
```
graph TD; dot(( )) --- the[the]; the --- sun[sun]; was[was] --- shining[shining]; her[her] --- hair[hair]; if[if] --- she[she]; she --- changed[changed]
```

Early one morning the sun

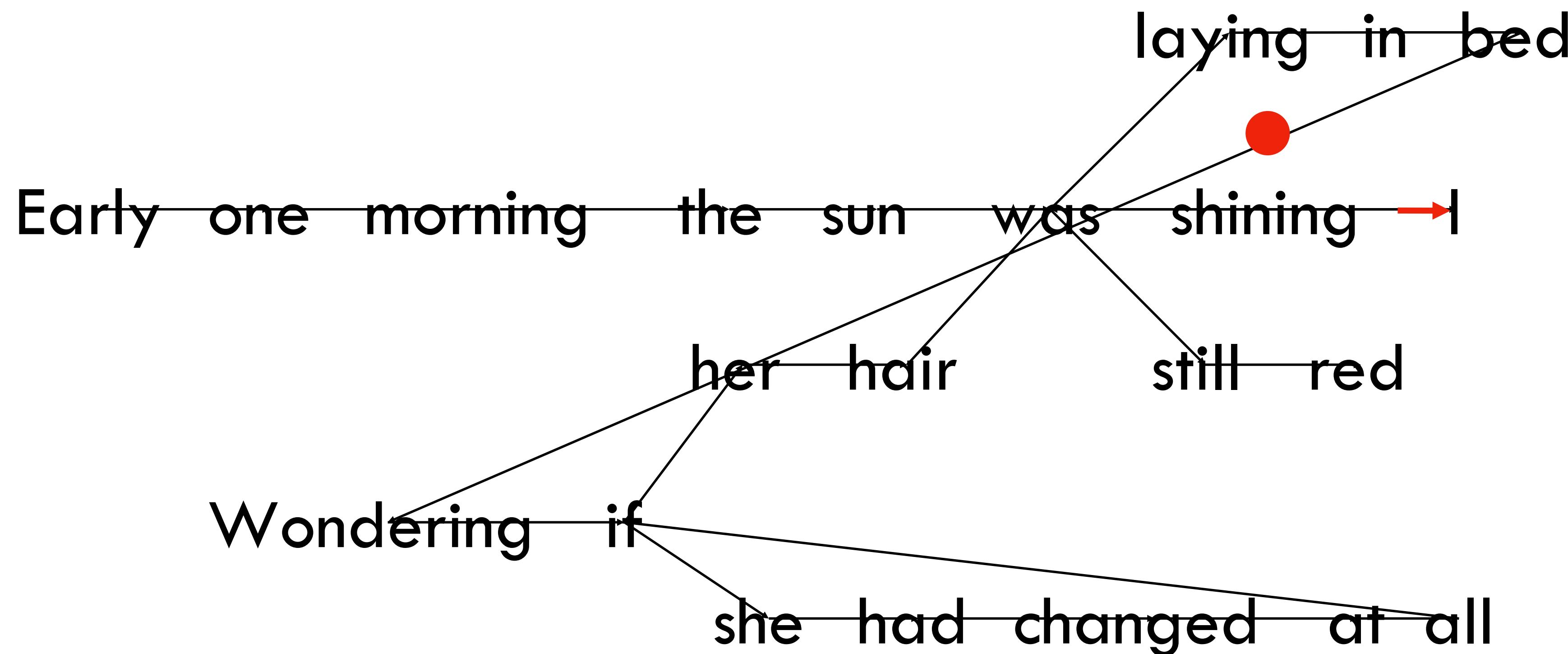
Early one morning the sun was shining I
laying in bed
her hair still red
Wondering if she had changed at all

```
graph TD; sun(( )) --- the[the]; sun --- sun[the sun]; was[was] --- shining[shining]; her[her] --- hair[hair]; she[she] --- changed[had changed at all]; changed --- changed[changed]
```

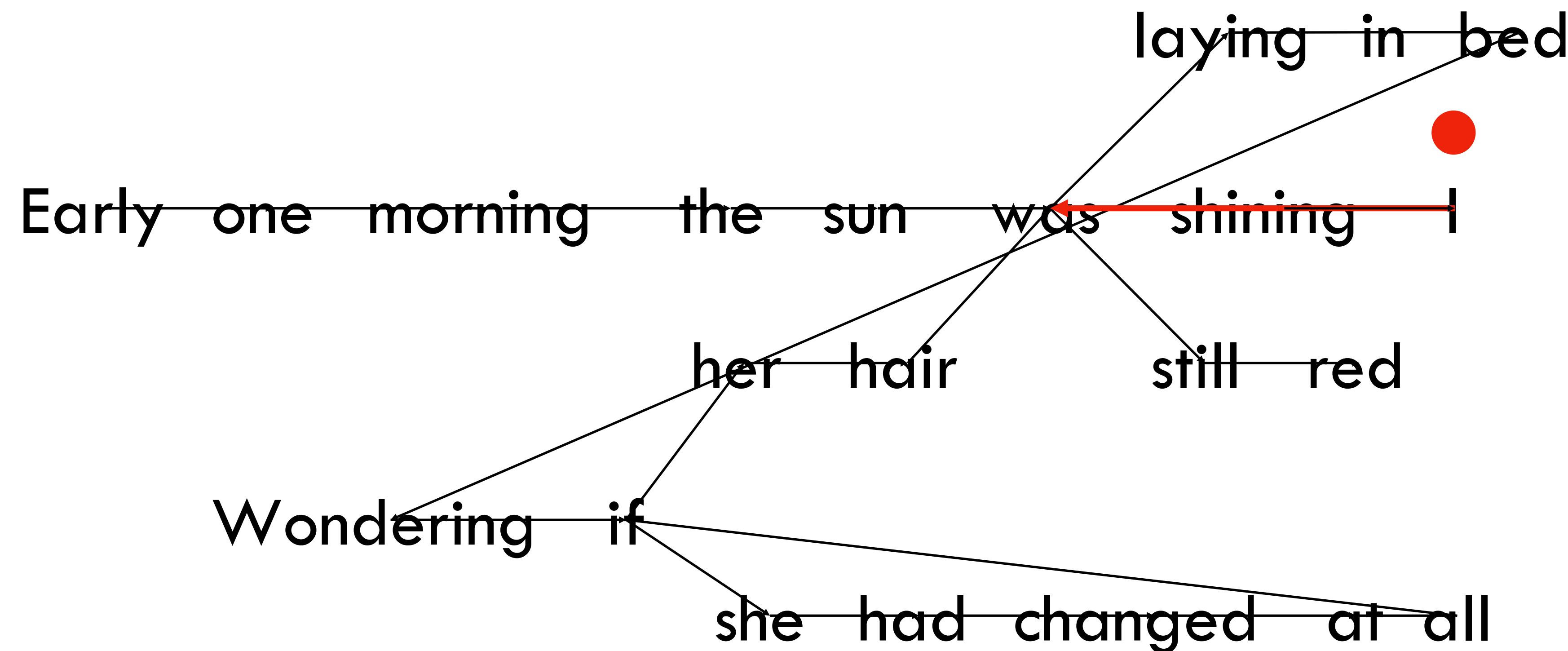
Early one morning the sun was



Early one morning the sun was shining



Early one morning the sun was shining |



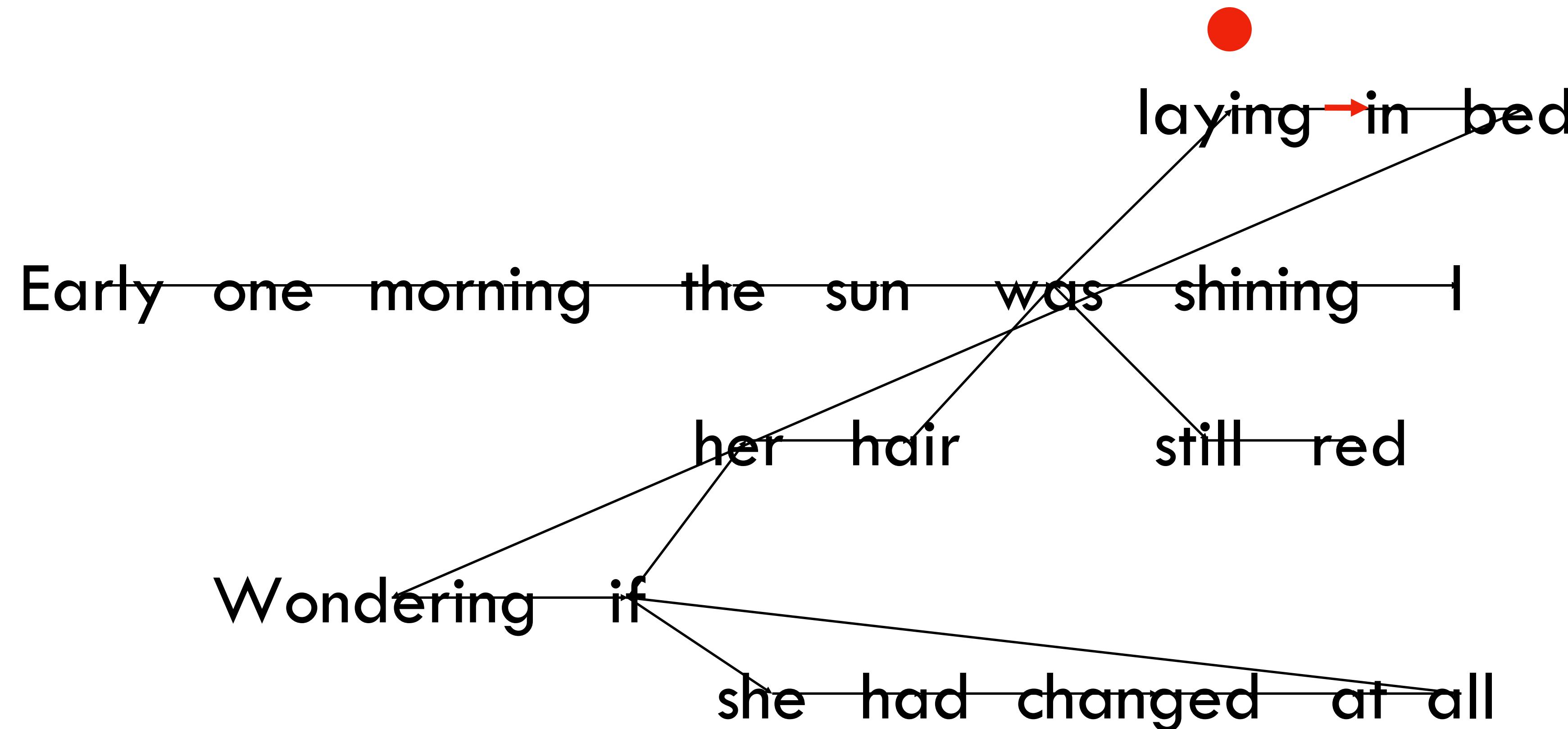
Early one morning the sun was shining I was

Early one morning the sun was shining I
was laying in bed

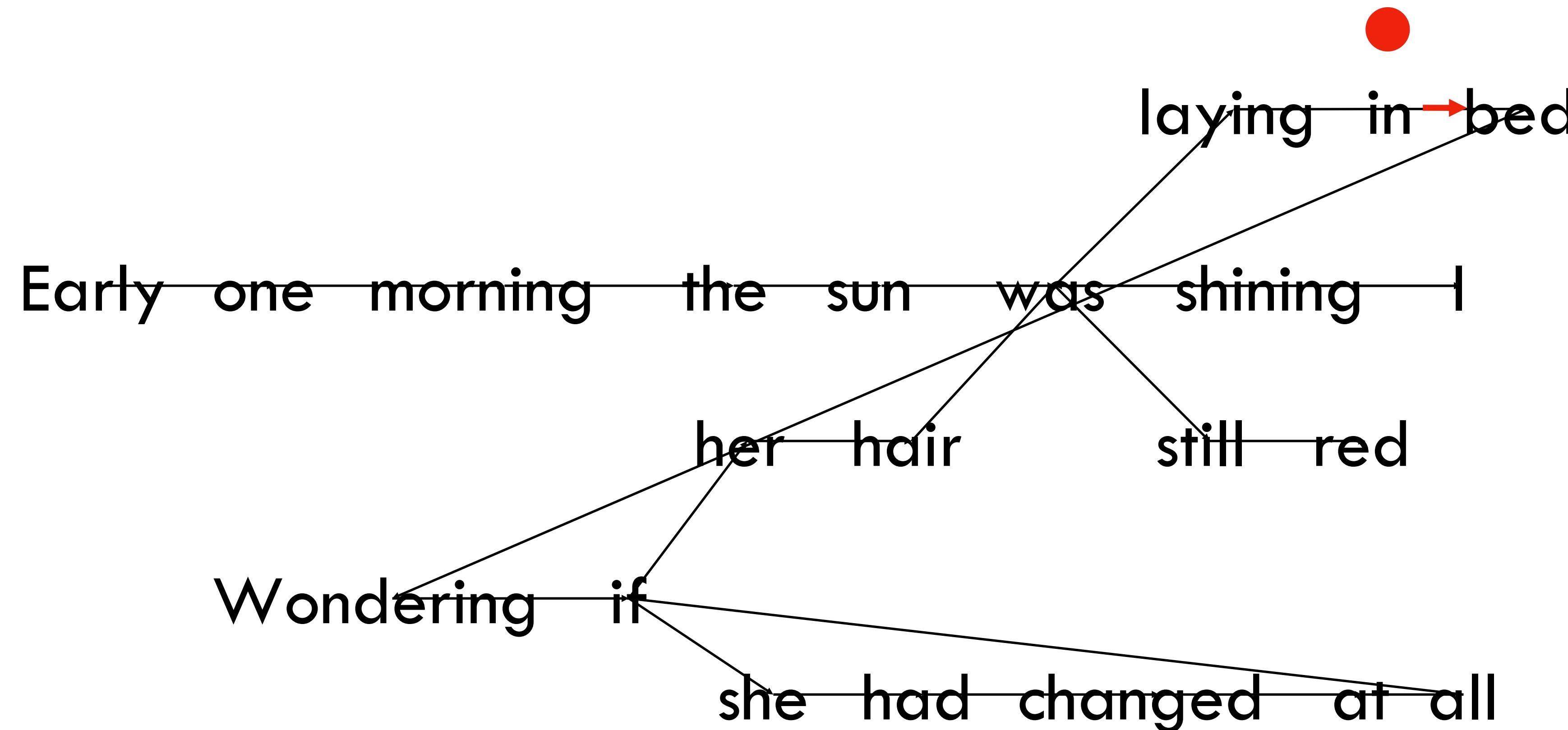
Wondering if her hair was still red
she had changed at all

```
graph TD; was((was)) --- laying[laying in bed]; was --- hair[her hair]; was --- changed[she had changed at all]; was --- I[ ]; was --- wonder[Wondering if]; was --- if;if --- hair; was --- she[ ]; she --- changed;
```

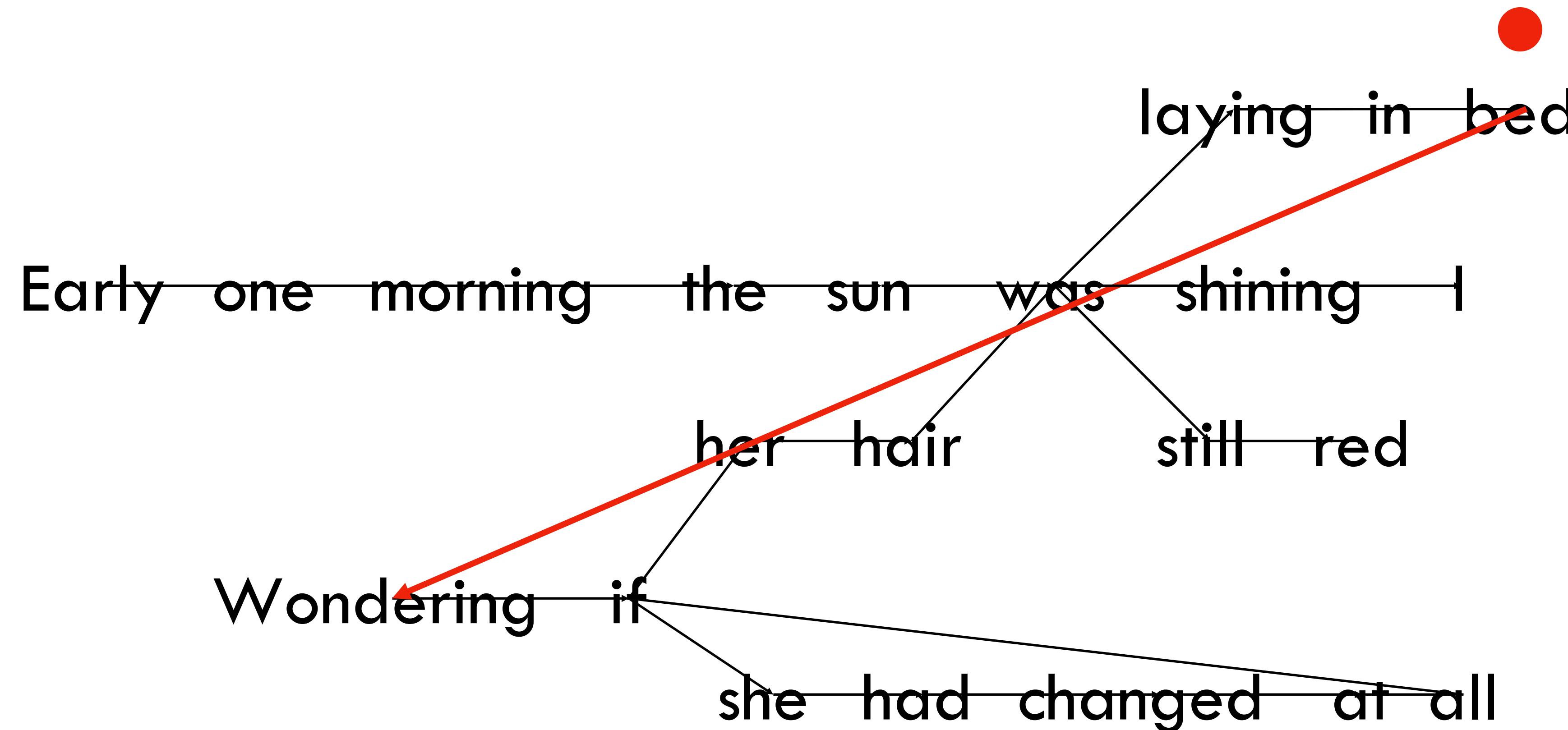
Early one morning the sun was shining I was laying



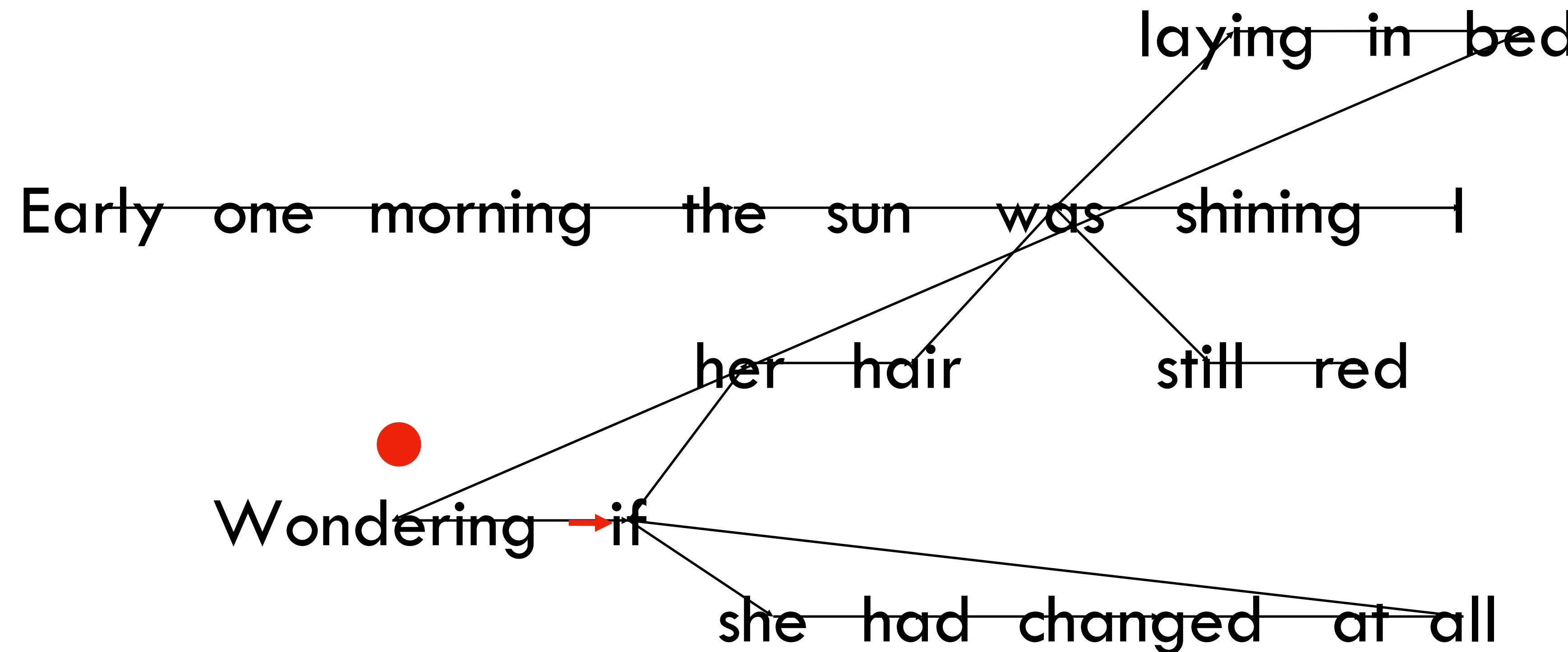
Early one morning the sun was shining I was laying in



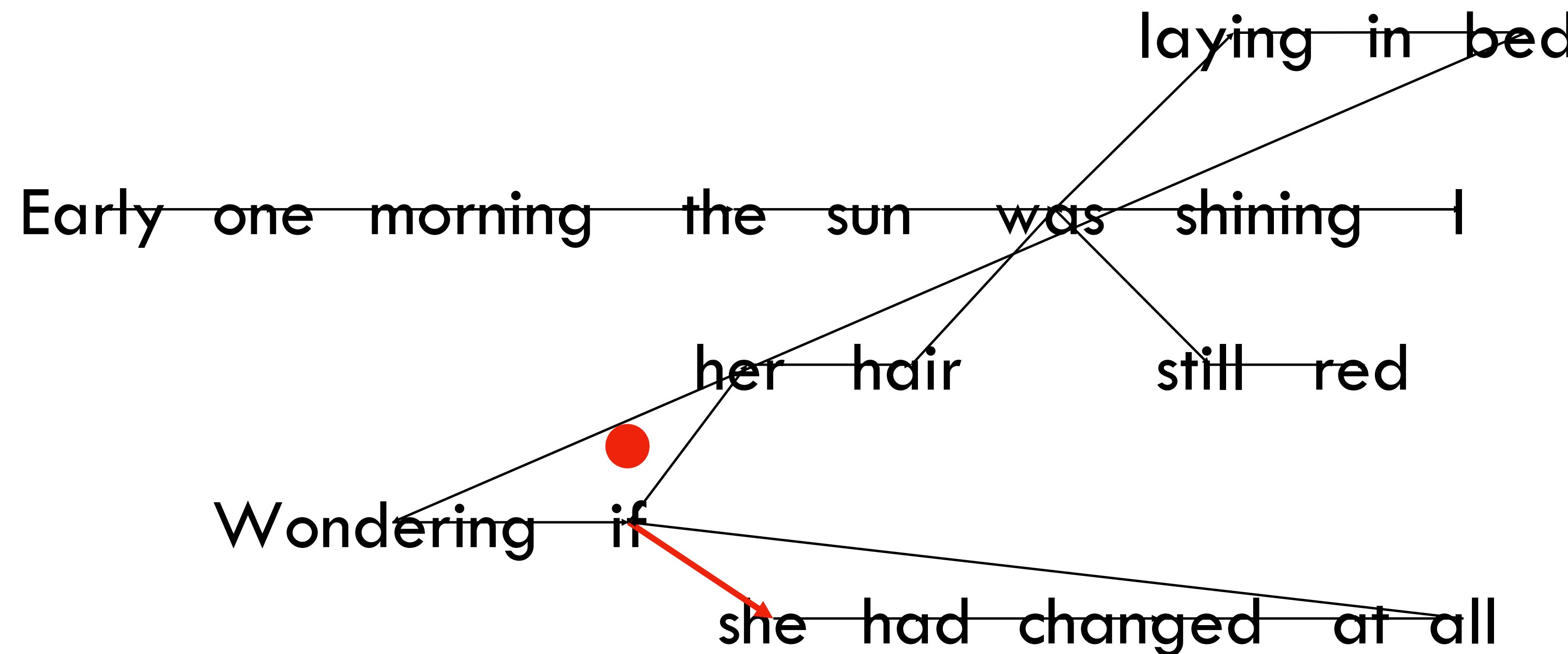
Early one morning the sun was shining I was laying in bed



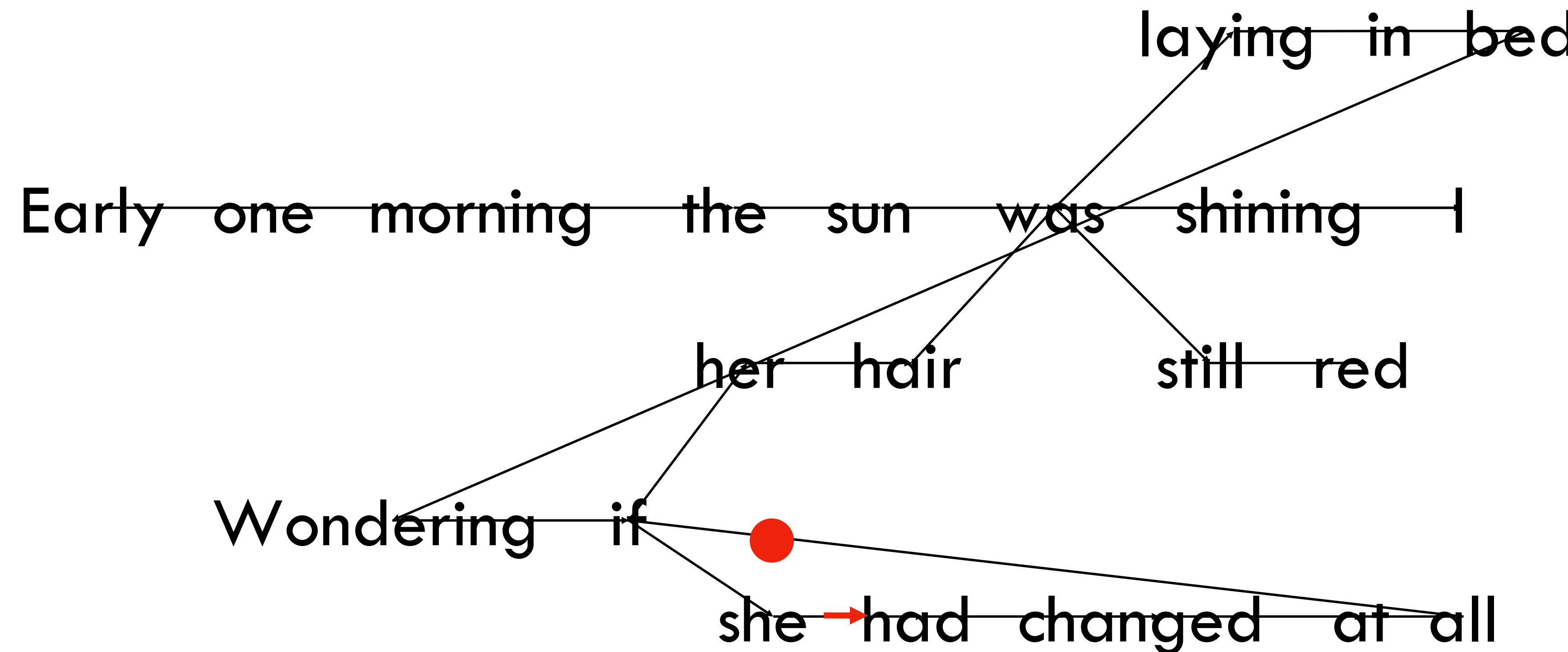
Early one morning the sun was shining I was laying in bed
Wondering



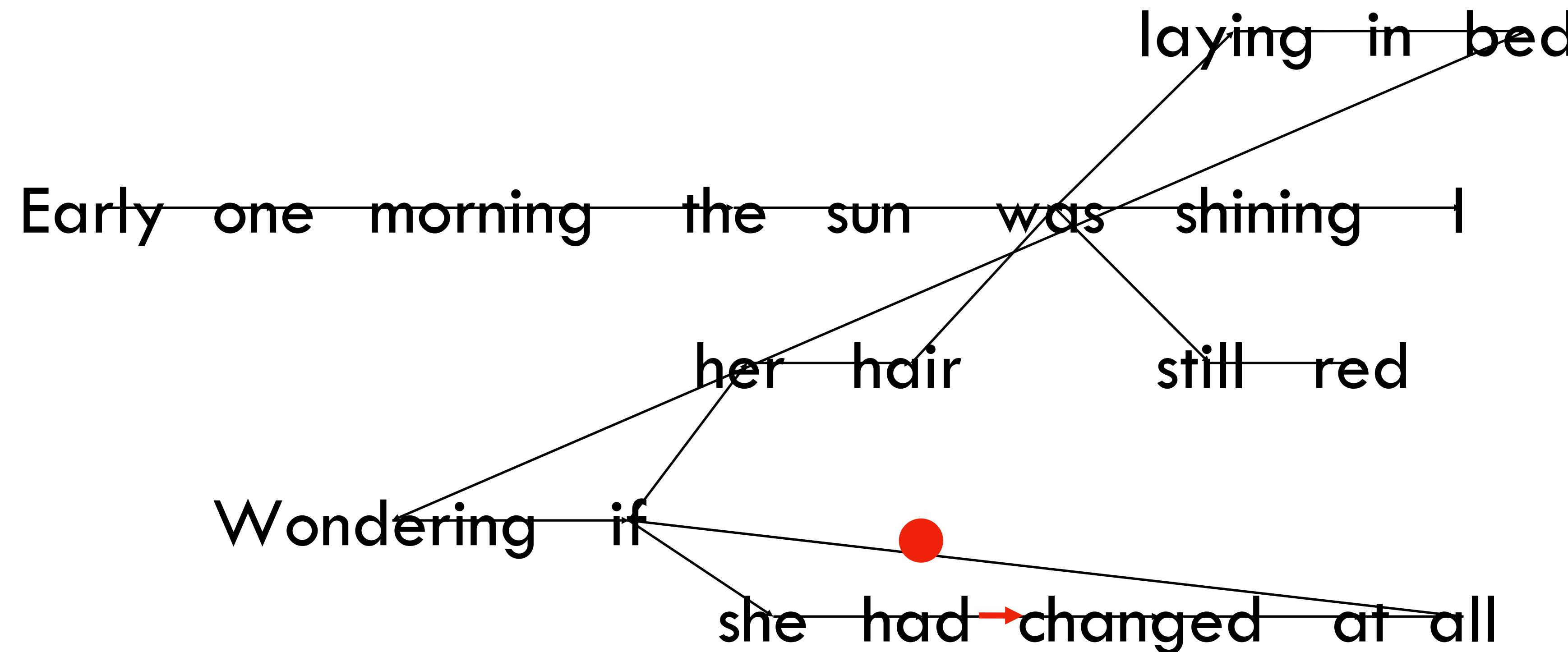
Early one morning the sun was shining I was laying in bed
Wondering if



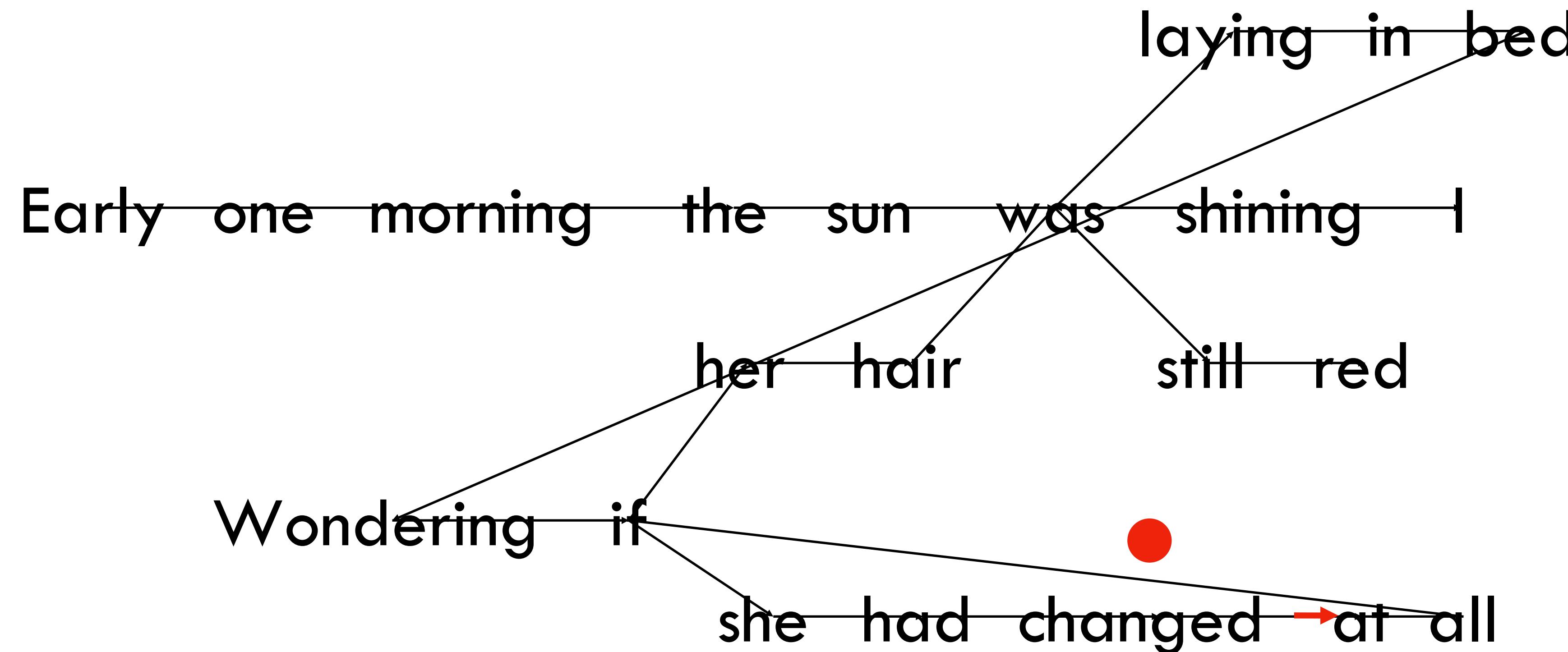
Early one morning the sun was shining I was laying in bed
Wondering if she



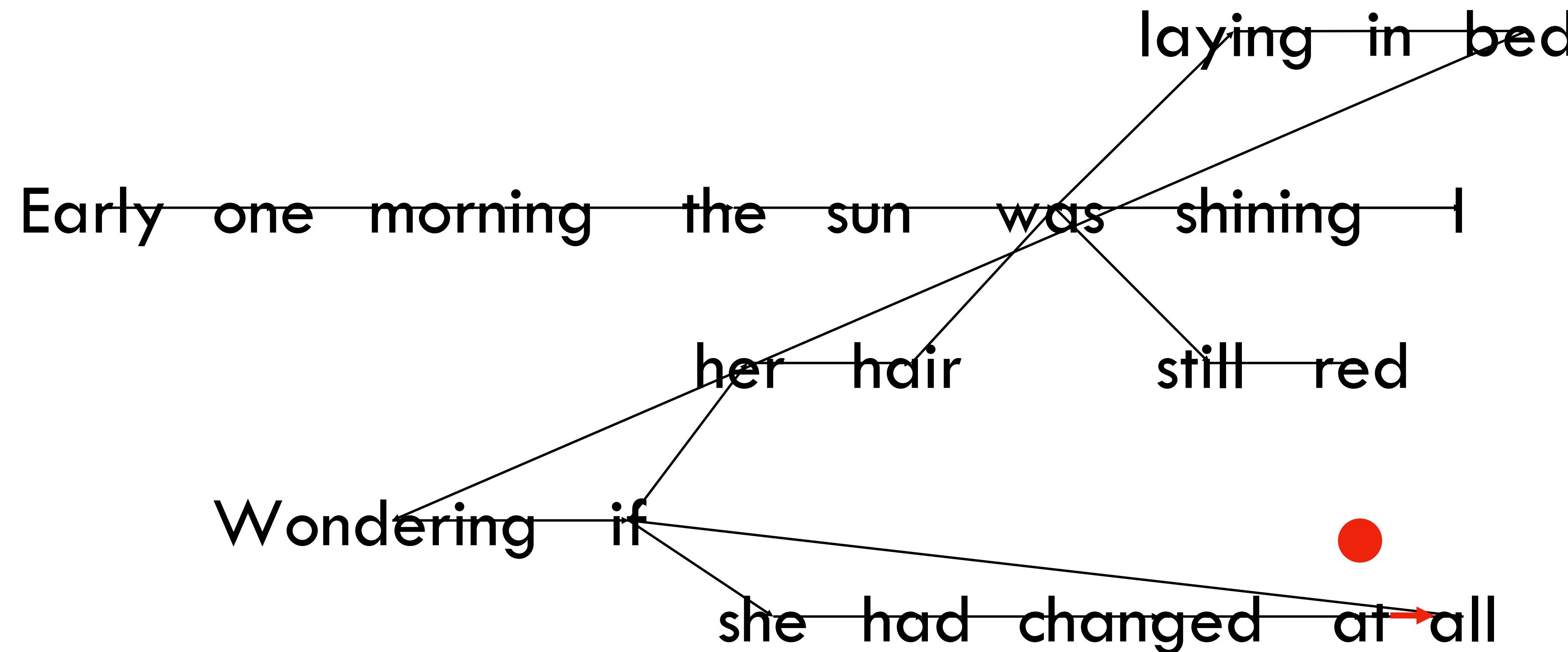
Early one morning the sun was shining I was laying in bed
Wondering if she had



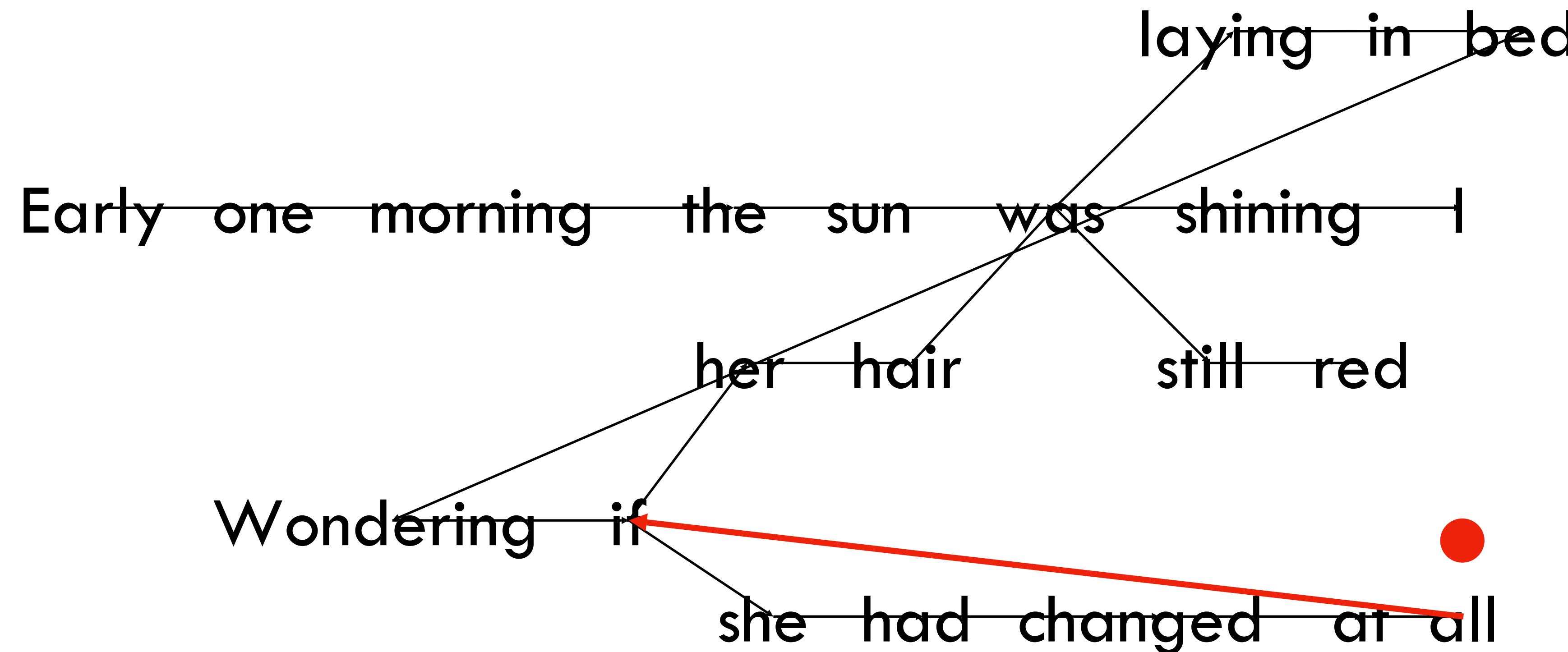
Early one morning the sun was shining I was laying in bed
Wondering if she had changed



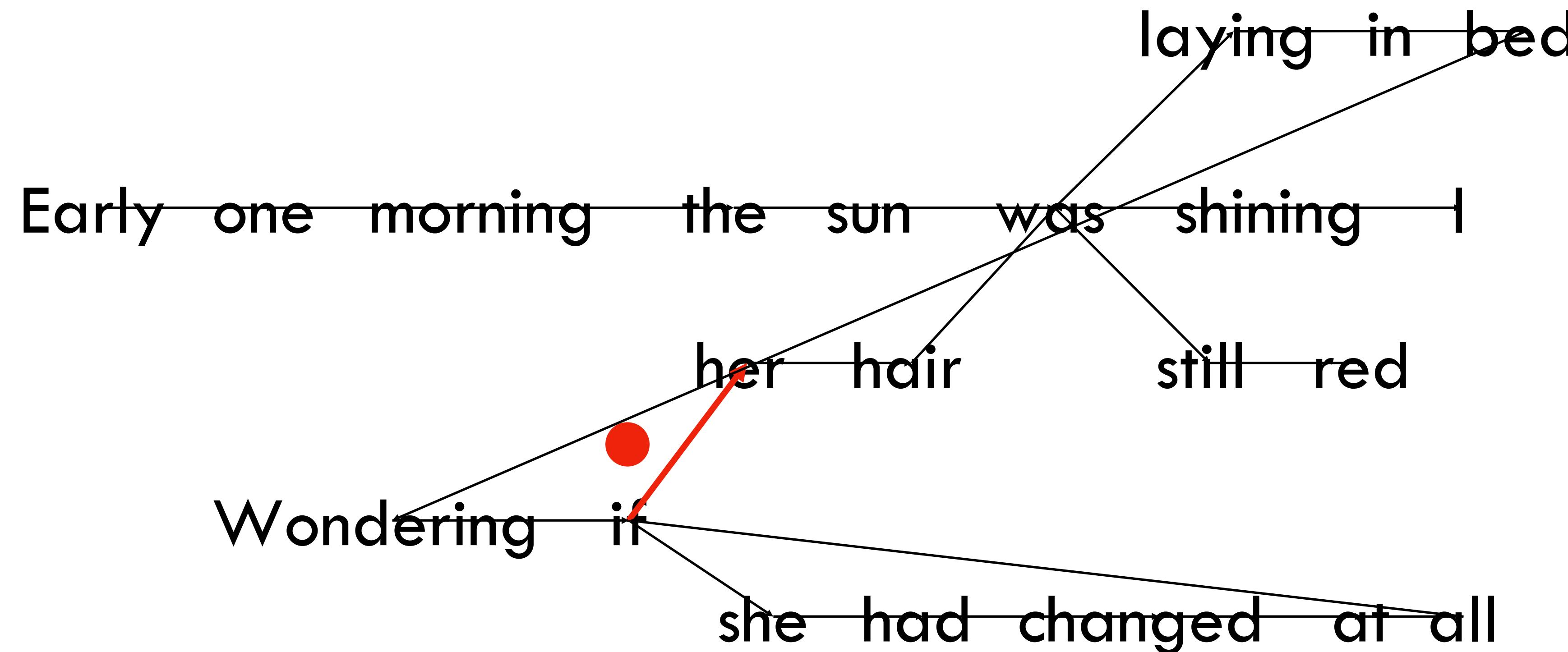
Early one morning the sun was shining I was laying in bed
Wondering if she had changed at



Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all

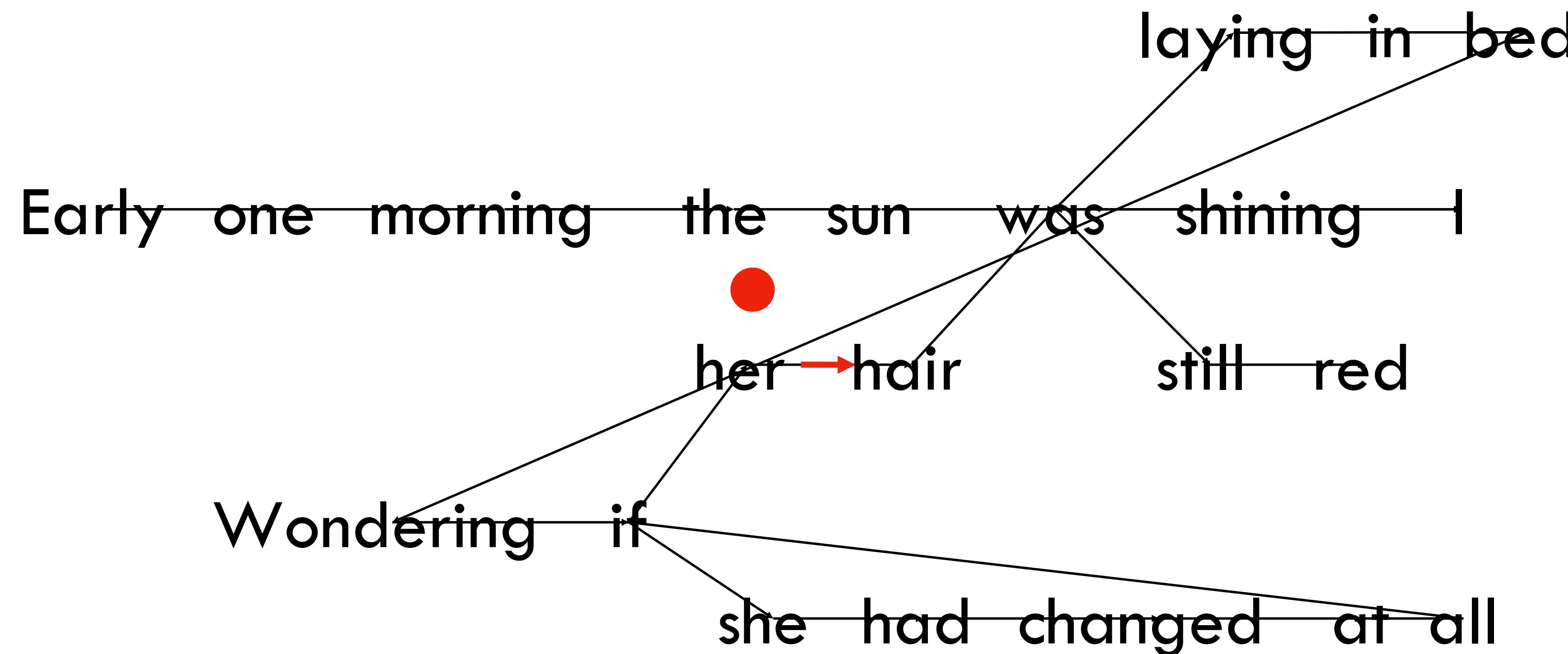


Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if

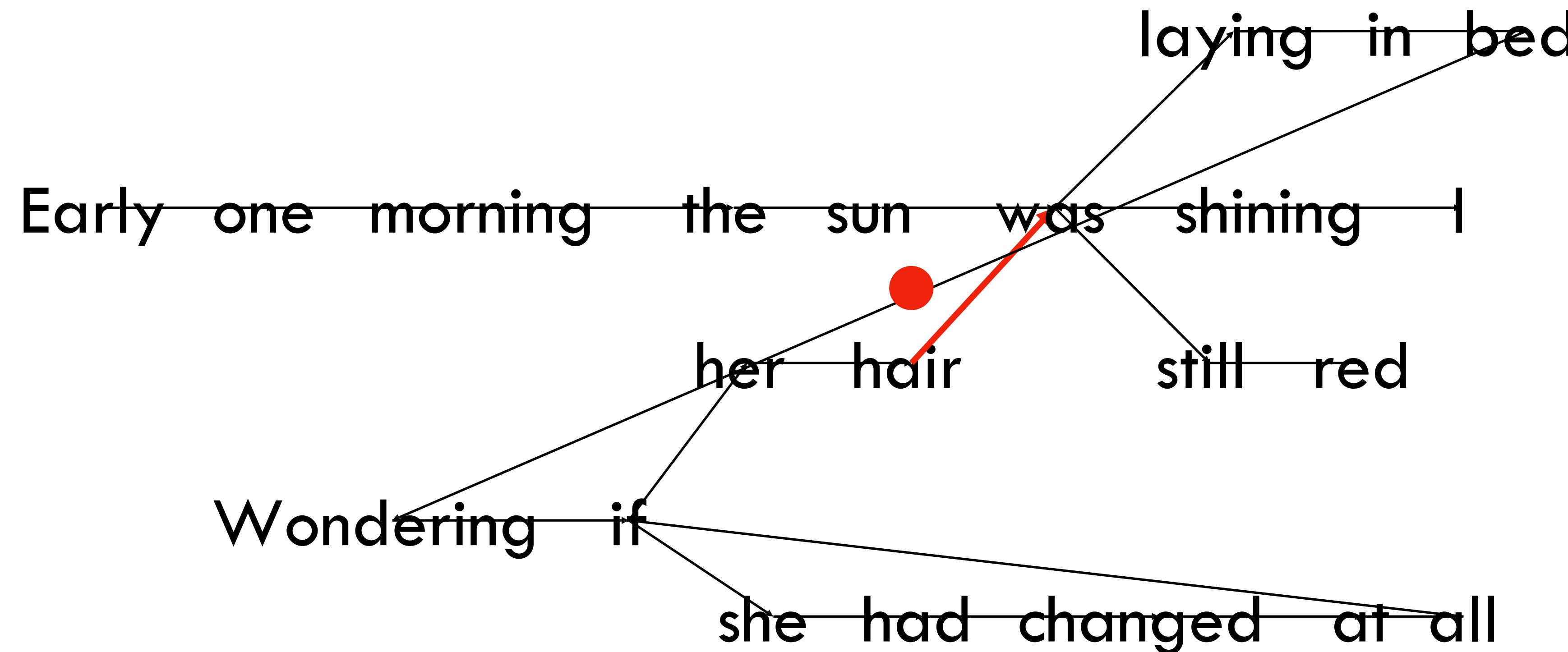


Early one morning the sun was shining I was laying in bed

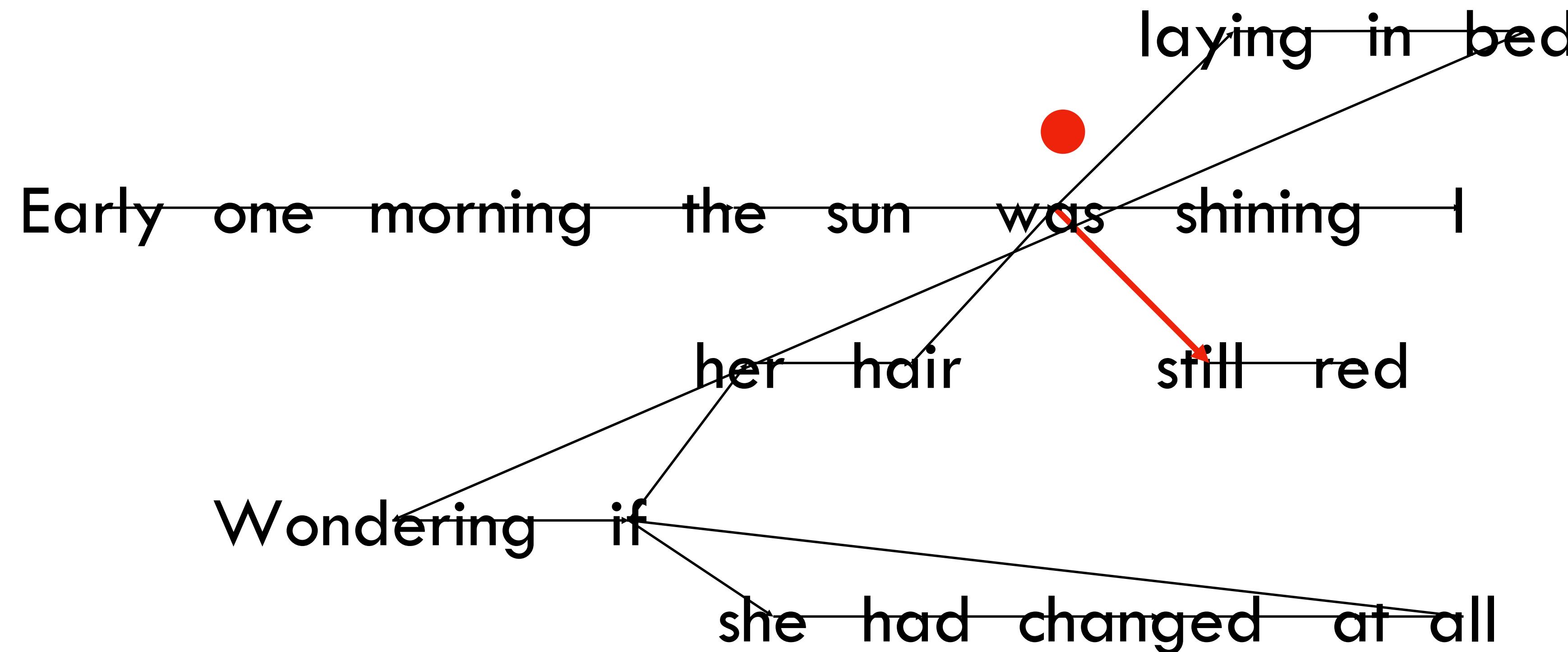
Wondering if she had changed at all if her



Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair

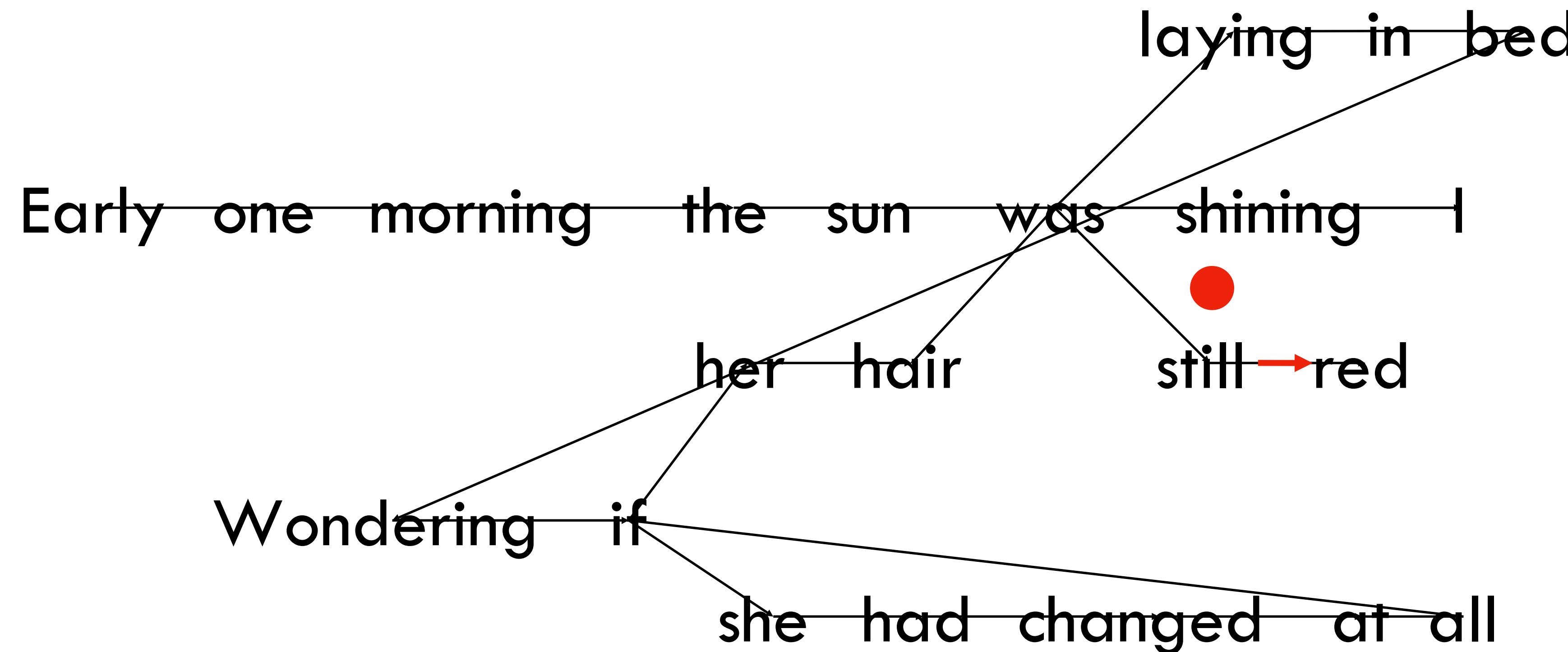


Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was

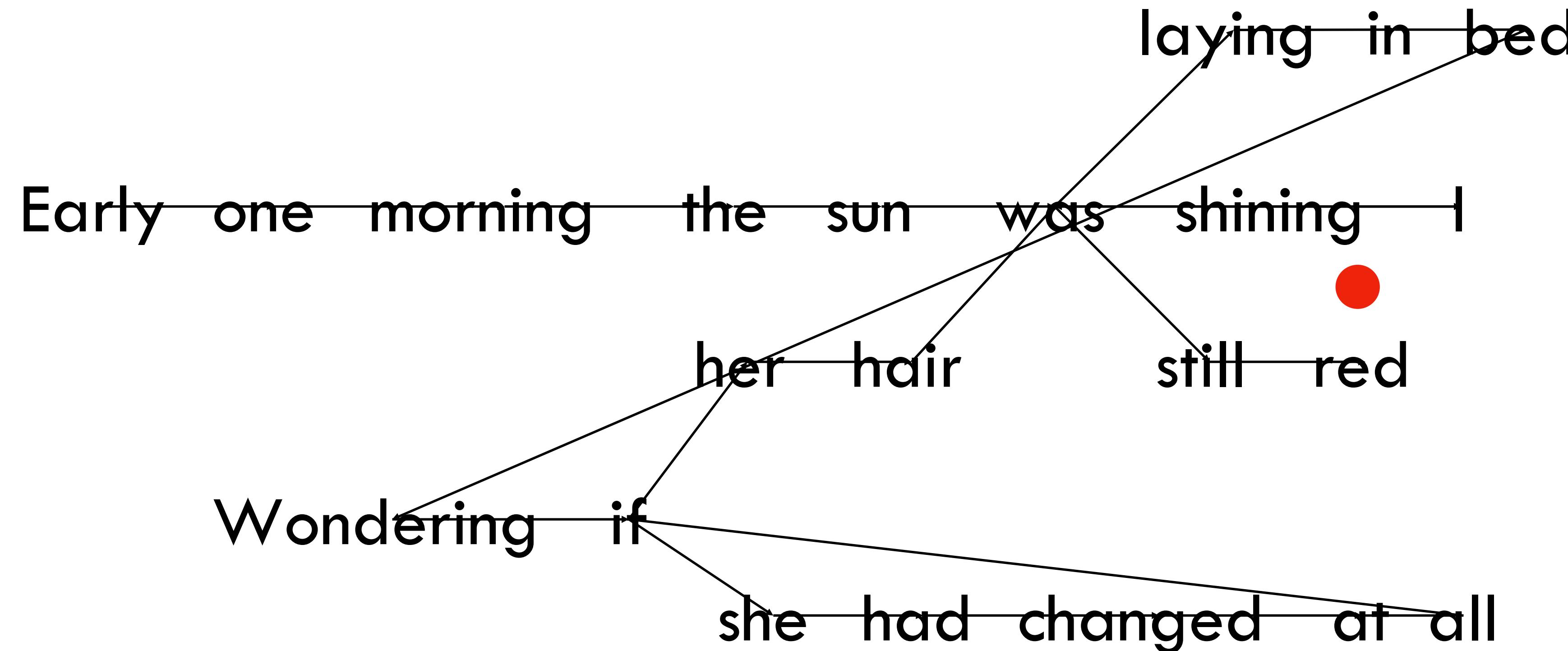


Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still



Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was still red



the

Early one morning the sun was shining I
laying in bed

Wondering if her hair still red

she had changed at all

```
graph TD; the((the)) --> sun[sun]; the --> was[was]; the --> shining[shining]; her[her] --> hair[hair]; she[she] --> changed[changed]; if[if] --- her; if --- changed;
```

the sun

Early one morning the sun was shining I
laying in bed

Wondering if her hair still red
she had changed at all

```
graph TD; Sun((sun)) --- Was[was]; Sun --- Her[her]; Sun --- She[she]; Was --- Shining[shining]; Her --- Hair[hair]; She --- HadChanged[had changed]; I[I] --- Shining
```

the sun was

Early one morning the sun was shining I
laying in bed

Wondering if her hair still red
she had changed at all

```
graph TD; was((was)) --- shining((shining)); was --- hair((hair)); was --- she((she)); was --- had((had)); if((if)) --- she; if --- had;
```

the sun was still

Early one morning the sun was shining I
laying in bed

Wondering if her hair still red
she had changed at all

```
graph LR; Early[Early] --- morning[morning]; sun[sun] --- shining[shining]; Wondering[Wondering] --- if;if --- her[her]; she[she] --- hair[hair]; had[had] --- changed[changed]; changed --- at[at]; changed --- all[all]; style dot fill:red,stroke:red,stroke-width:2px; dot((still))
```

the sun was still red

Early one morning the sun was shining I
laying in bed

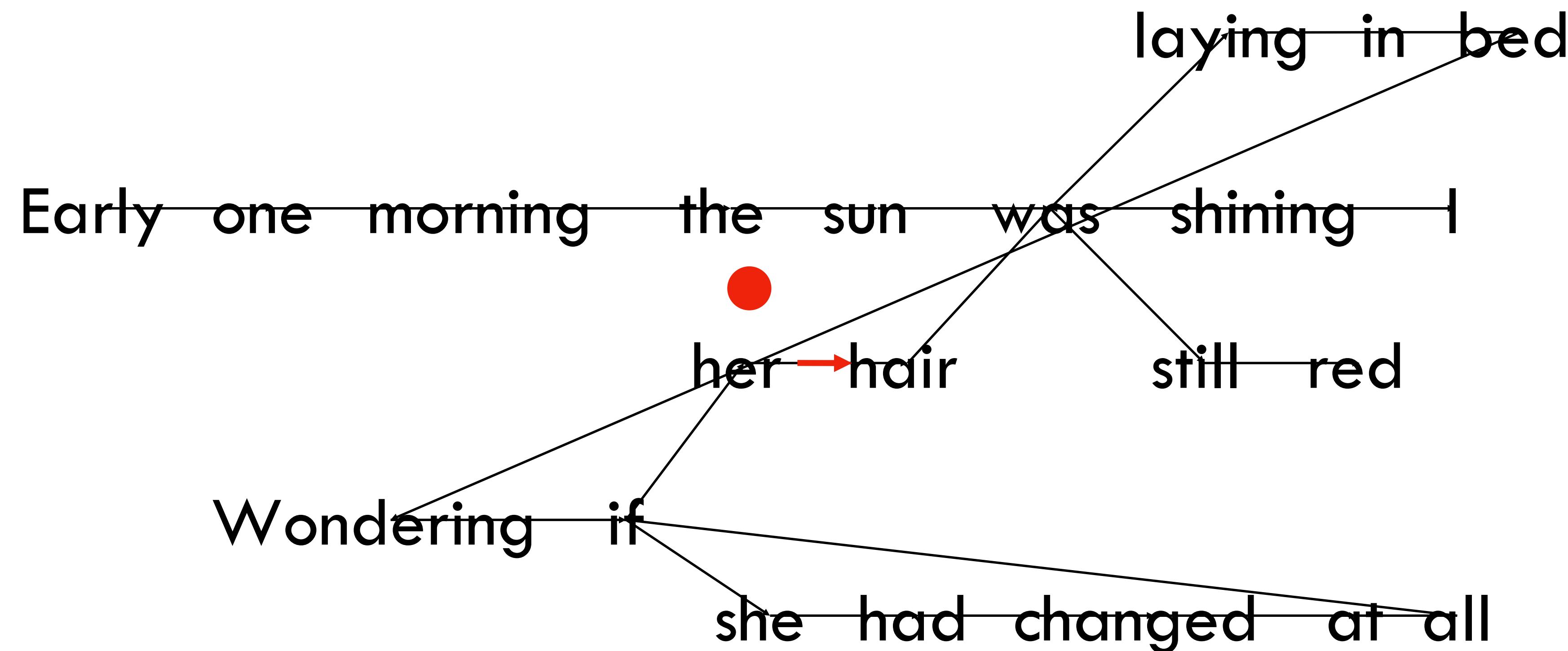
Wondering if her hair still red

she had changed at all

The diagram illustrates a semantic network or a conceptual graph. The top line of text ('Early one morning the sun was shining I laying in bed') is connected by lines to the bottom line ('Wondering if her hair still red') and the bottom-most line ('she had changed at all'). Specifically, 'sun' connects to 'her hair', 'was' connects to 'still', and 'shining I laying in bed' connect to 'Wondering if' and 'she had changed at all'. A single red dot is positioned above the word 'red'.

the sun was still red

her



the sun was still red

her hair

Early one morning the sun was shining I
laying in bed

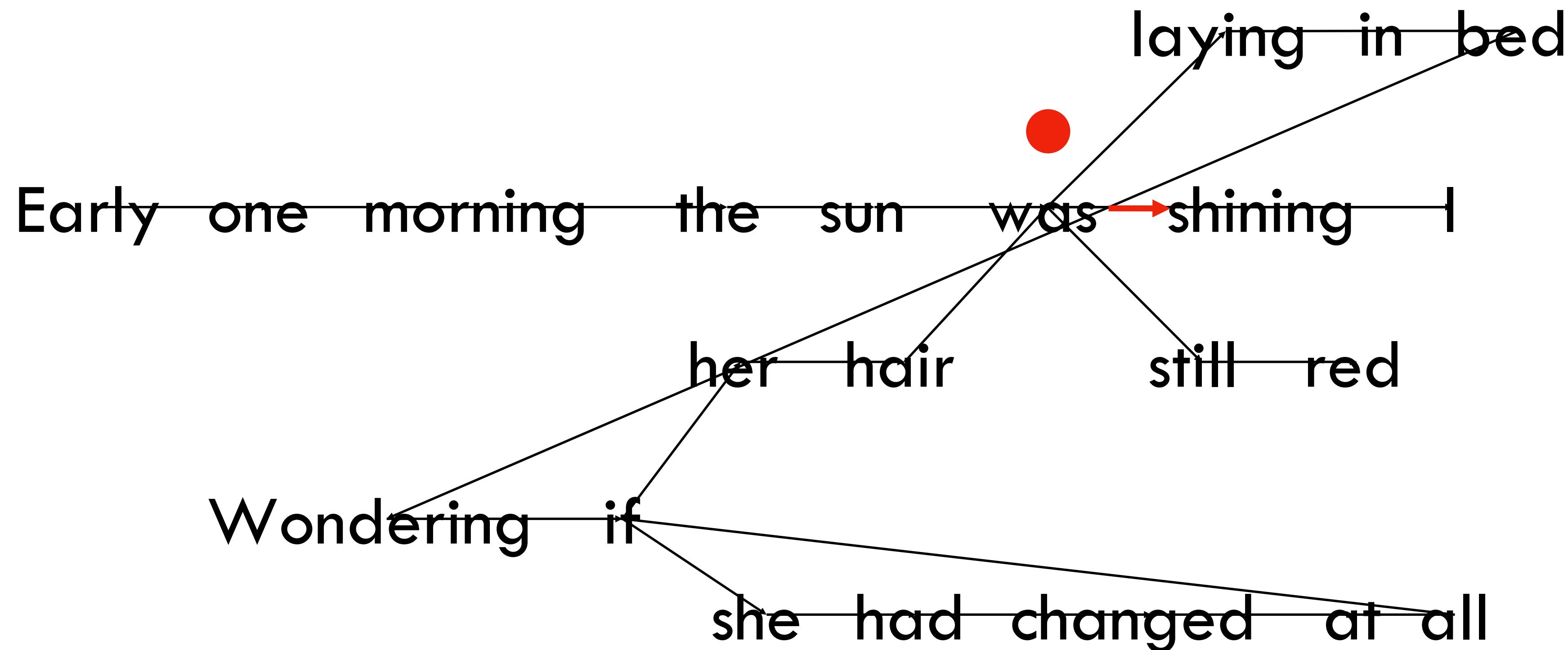
Wondering if her hair was still red

she had changed at all

```
graph TD; A(( )) --- B(( )); C(( )) --- D(( )); E(( )) --- F(( )); G(( )) --- H(( )); I(( )) --- J(( )); K(( )) --- L(( )); M(( )) --- N(( )); O(( )) --- P(( )); Q(( )) --- R(( )); S(( )) --- T(( )); U(( )) --- V(( )); W(( )) --- X(( )); Y(( )) --- Z(( ));
```

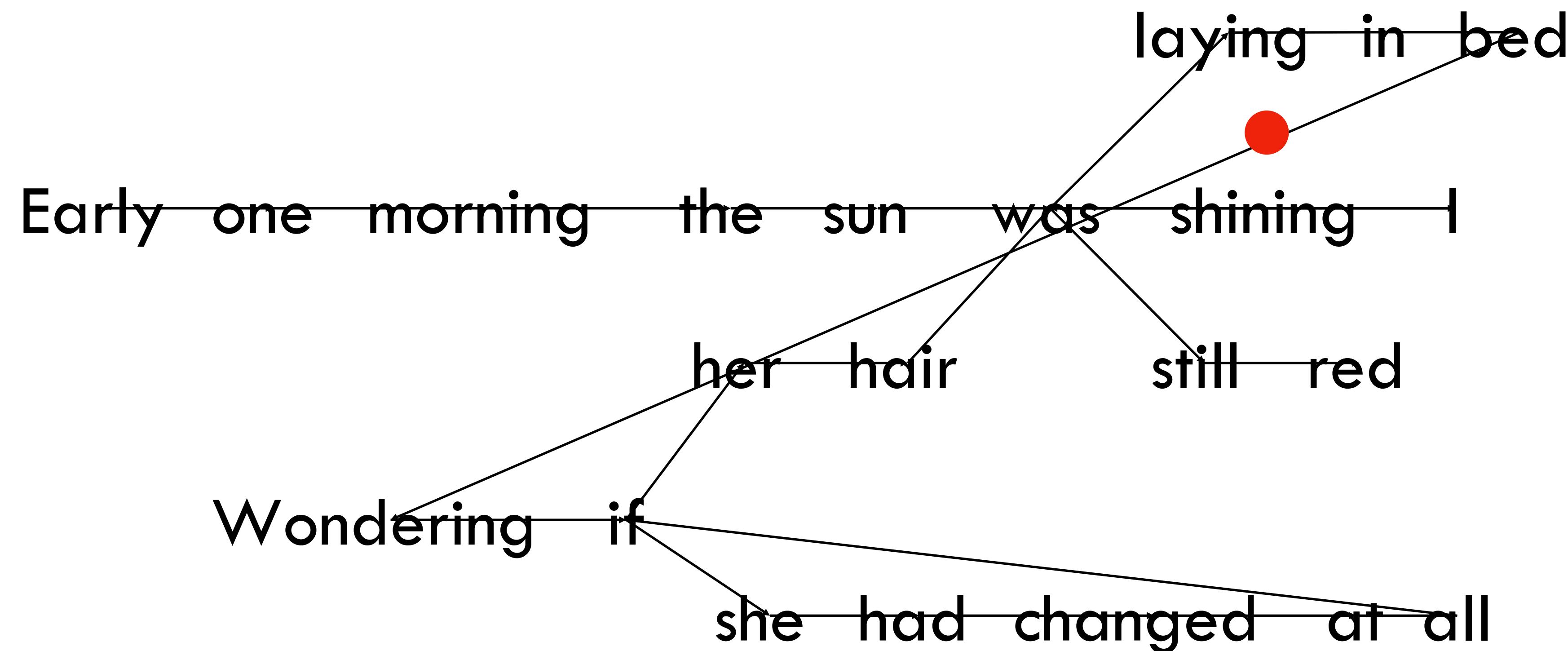
the sun was still red

her hair was



the sun was still red

her hair was shining



the

Early one morning the sun was shining I
laying in bed

Wondering if her hair still red

she had changed at all

```
graph TD; the((the)) --> sun(sun); the --> was(was); the --> shining(shining); her --> hair(hair); she --> changed(changed); if(if) --- her; if --- changed; at(at) --- all(all)
```

the sun

Early one morning the sun was shining I
laying in bed

Wondering if her hair still red
she had changed at all

```
graph TD; Sun((sun)) --- Was[was]; Sun --- Her[her]; Sun --- She[she]; Was --- Shining[shining]; Her --- Hair[hair]; She --- HadChanged[had changed]; I[I] --- Shining
```

the sun was

Early one morning the sun was shining I
laying in bed

Wondering if her hair still red
she had changed at all

```
graph TD; Sun((Sun)) --- Was[was]; Was --- Shining[shining]; Was --- Bed[bed]; Her[her] --- Hair[hair]; If;if --- Changed[changed];
```

the sun was laying

Early one morning the sun was shining |
Wondering if her hair still red
she had changed at all

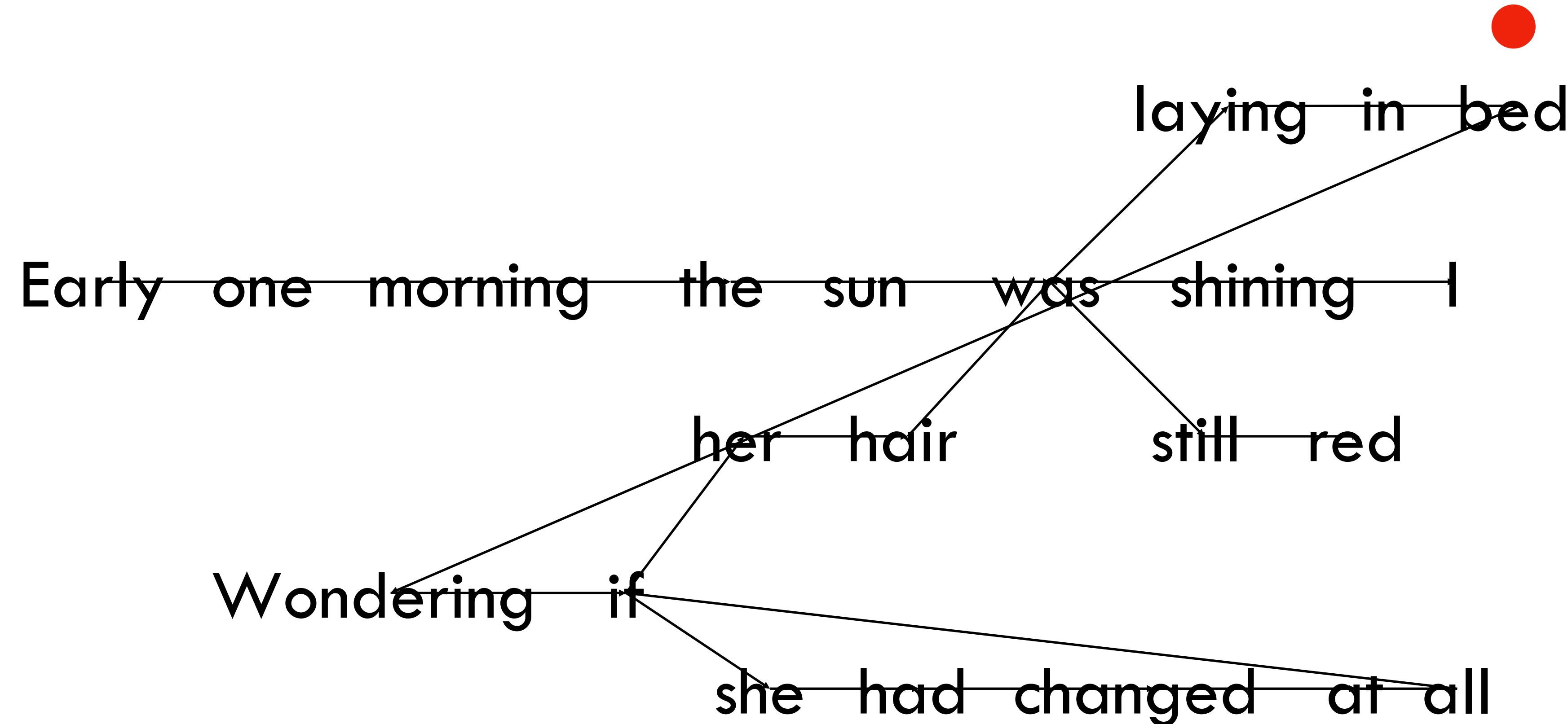
```
graph TD; dot(( )) --- laying[laying]; was[was] --- shining[shining]; her[her] --- hair[hair]; if(if) --- she[she]; if(if) --- had[had];
```

the sun was laying in

Early one morning the sun was shining |
Wondering if her hair still red
she had changed at all

The diagram illustrates the semantic relationships between words in the sentence. It features a horizontal baseline with vertical lines extending upwards from each word. A red dot is positioned above the word 'sun'. Lines connect 'sun' to 'shining', 'her' to 'hair', and 'she' to 'changed'. Additionally, lines connect 'Wondering' to 'if', 'if' to 'her', and 'if' to 'she'.

the sun was laying in bed



Early one morning the sun was shining laying in bed

Wondering if her hair was still red

she had changed at all

I was

Early one morning the sun was shining I
laying in bed

Wondering if her hair still red
she had changed at all

The diagram illustrates a sentence structure with several words highlighted in black. Arrows point from the word 'was' to the word 'shining' and from the word 'if' to the word 'changed'. A red dot is placed above the word 'was', and a red arrow points from it to the word 'shining'.

I was shining

Early one morning the sun was shining I

her hair still red

Wondering if she had changed at all

I was shining I

Early one morning the sun was shining
Wondering if her hair still red
laying in bed

```
graph TD; Early[Early one morning] --- sun[the sun]; sun --- was[was]; was --- shining[shining]; Wondering[Wondering if] --- her[her]; her --- hair[hair]; she[she] --- had[had changed at all]; had --- at[at all]; laying[laying in bed]; laying --- bed[bed]
```

I was shining I was

Early one morning the sun was shining I
was laying in bed

Wondering if her hair was still red

she had changed at all

```
graph TD; Early[Early] --- one[one]; one --- morning[morning]; morning --- the[the]; the --- sun[sun]; sun --- was1[was]; was1 --- shining[shining]; shining --- I[I]; Wondering[Wondering] --- if;if --- her[her]; her --- hair[hair]; hair --- was2[was]; was2 --- still[still]; still --- red[red]; she[she] --- had[had]; had --- changed[changed]; changed --- at[at]; I --- laying[laying]; laying --- in[in]; in --- bed[bed]; at --- all[all]
```

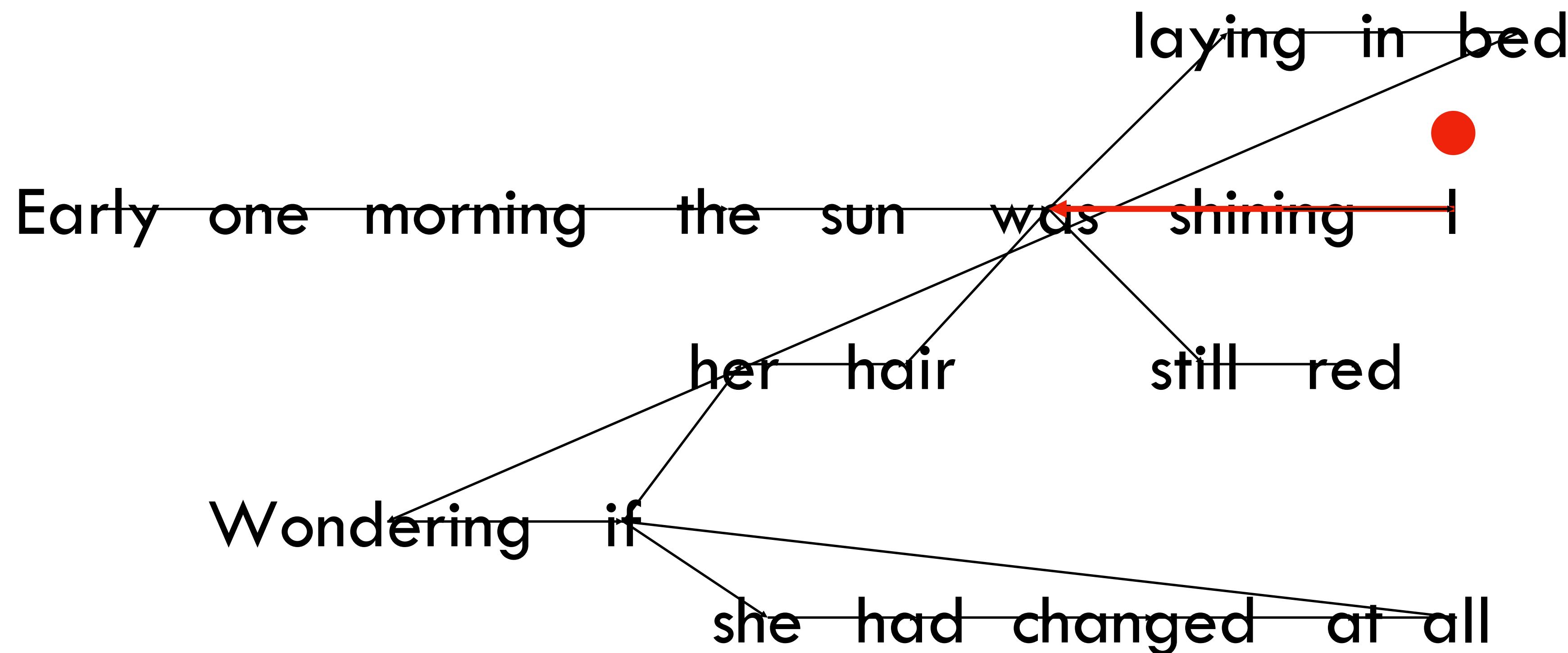
I was shining I was shining

Early one morning the sun was shining I

her hair still red

Wondering if she had changed at all

I was shining I was shining I



I was shining I was shining I was

Early one morning the sun was shining I
was laying in bed

Wondering if her hair was still red

she had changed at all

```
graph TD; was1[was] --- shining1[shining]; was1 --- red[red]; changed[she had changed] --- atall[at all]
```

I was shining I was shining I was still

Early one morning the sun was shining I
was laying in bed

Wondering if her hair was still red
she had changed at all

```
graph TD; A["I was shining I was shining I was still"] --> B["Early one morning the sun was shining I was laying in bed"]; B --> C["Wondering if her hair was still red she had changed at all"]; C --> D["red"]
```

I was shining I was shining I was still red

Early one morning the sun was shining I
laying in bed

Wondering if her hair still red

she had changed at all

```
graph TD; Early[Early] --- morning[morning]; the[the] --- sun[sun]; was[was] --- shining[shining]; I[I] --- bed[bed]; Wondering[Wondering] --- if;if --- her[her]; her --- hair[hair]; she[she] --- changed[changed]; at[at] --- all[all]; style redDot fill:red,stroke:red,stroke-width:2px;
```

$$P(x_n | x_{n-1})$$

Early one morning the sun was shining

Wondering if her hair was still red

she had changed at all

laying in bed

```
graph TD; was[was] --- shining[shining]; was --- hair[her hair]; was --- changed[changed]; if[if] --- changed;
```

Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

Early one morning

trigrams

Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

Early one morning
one morning the

trigrams

Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

Early one morning
one morning the
morning the sun

trigrams

Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

Early one morning
one morning the
morning the sun
the sun was

trigrams

Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was still red

Early one morning
one morning the
morning the sun
the sun was

trigrams

slide from Steve Seitz's [video](#)

Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was still red

Early one morning
one morning the
morning the sun
the sun was
sun was shining
was shining I
shining I was
I was laying
...

trigrams

$$P(x_n | x_{n-1}, x_{n-2})$$

Early one → morning
one morning → the
morning the → sun
the sun → was
sun was → shining
was shining → I
shining I → was
I was → laying
...

Video Textures

Arno Schödl

Richard Szeliski

David Salesin

Irfan Essa

Microsoft Research, Georgia Tech

SIGGRAPH 2000

Still photos



Video clips



Video textures



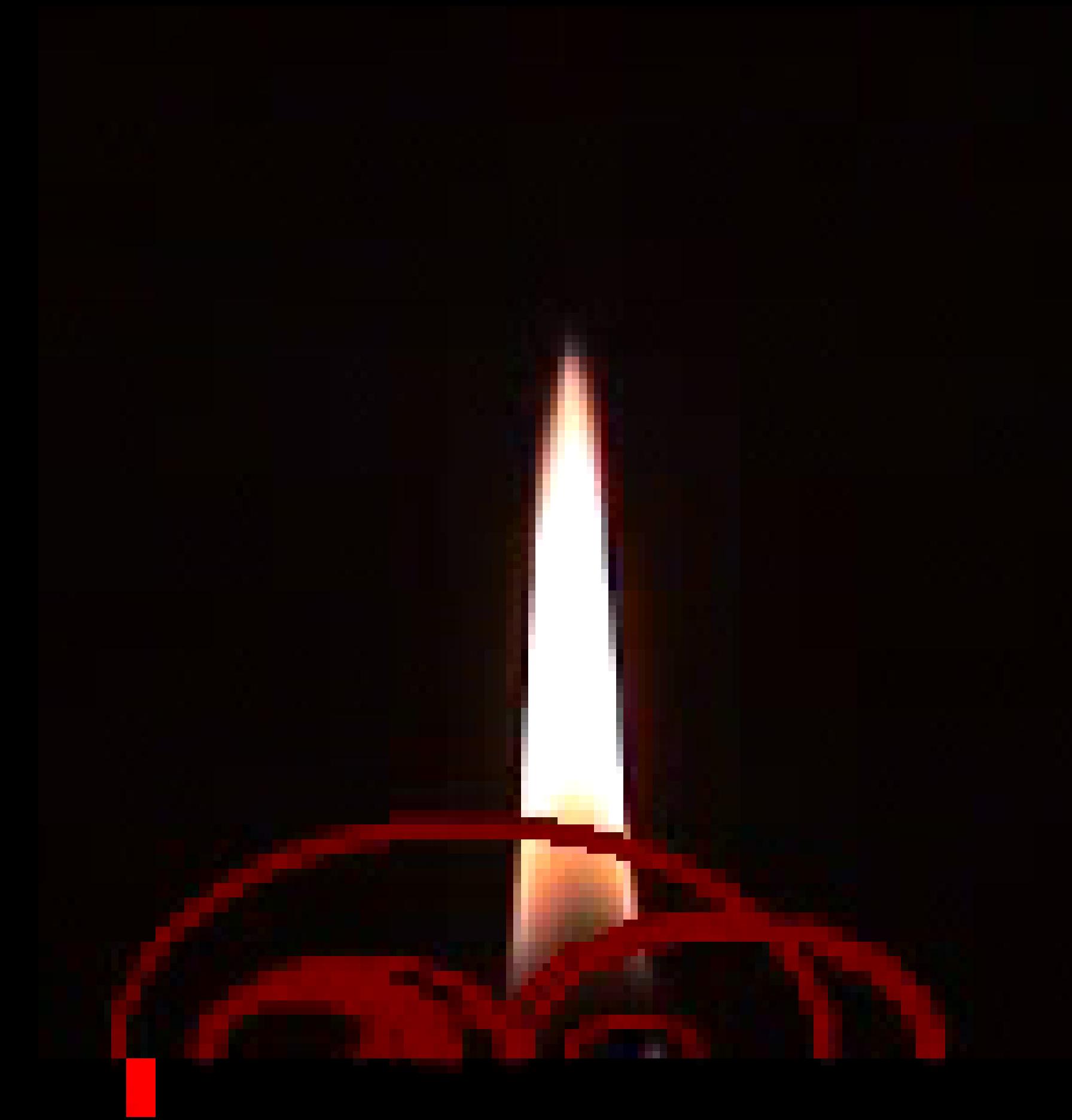
Problem statement



video clip

video texture

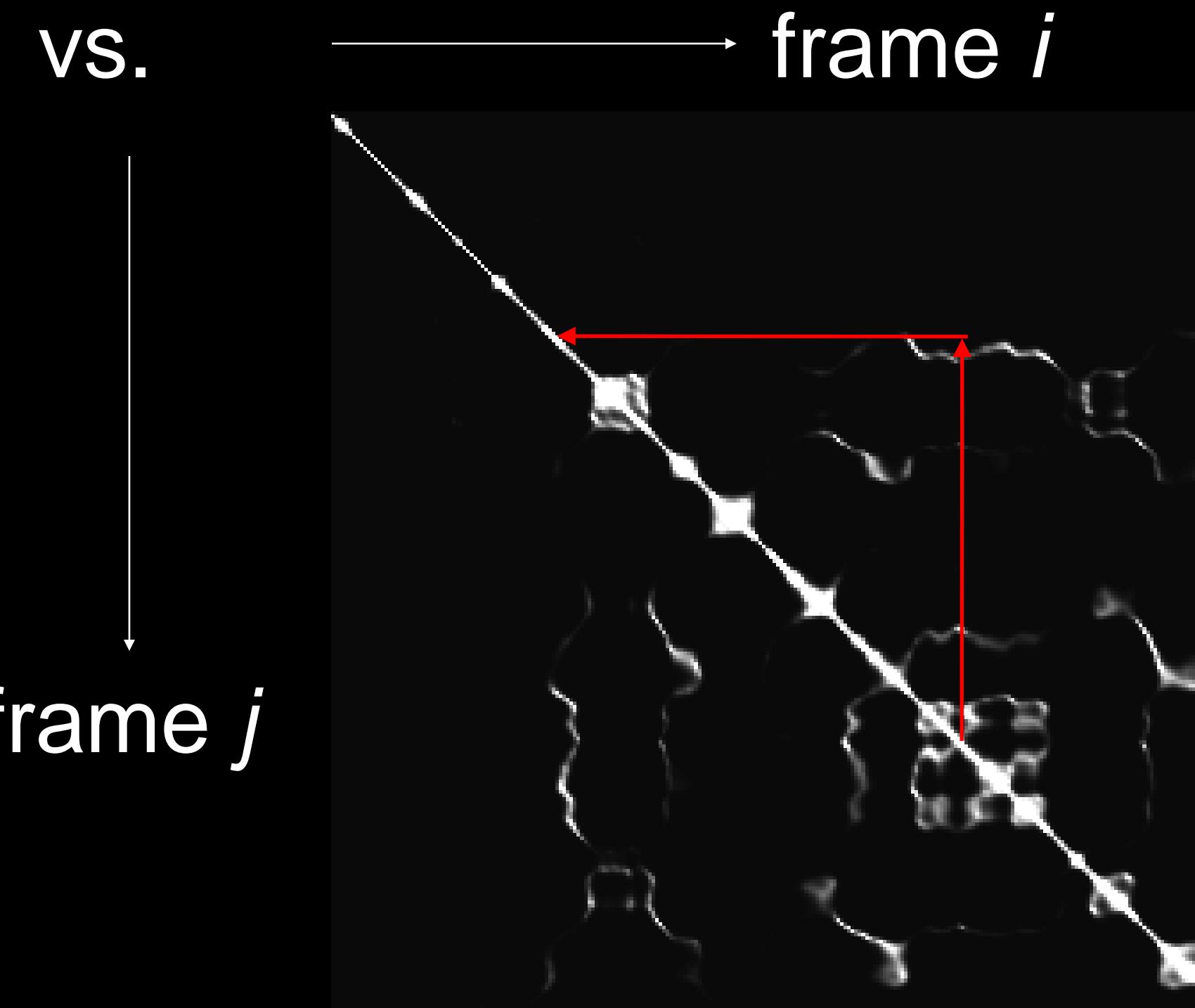
Our approach



- How do we find good transitions?

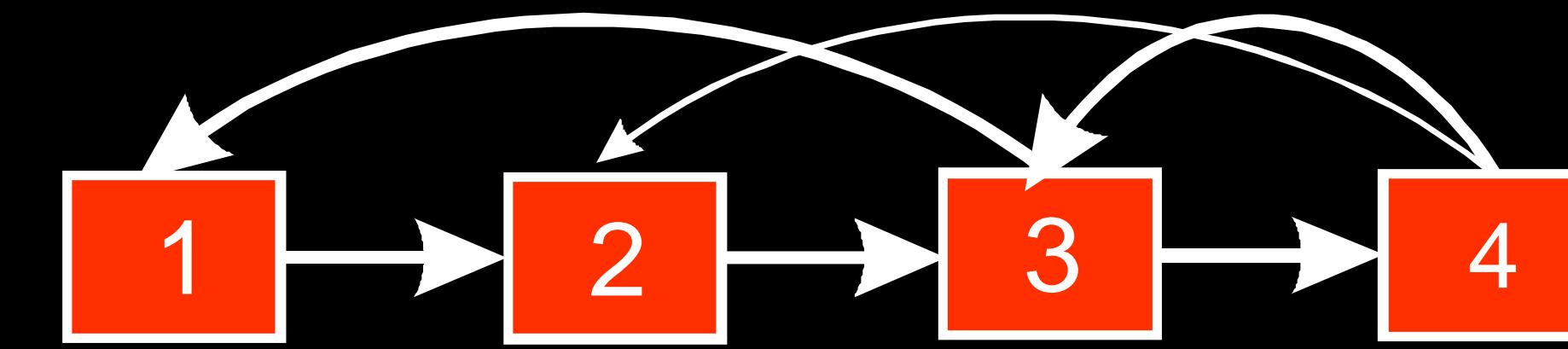
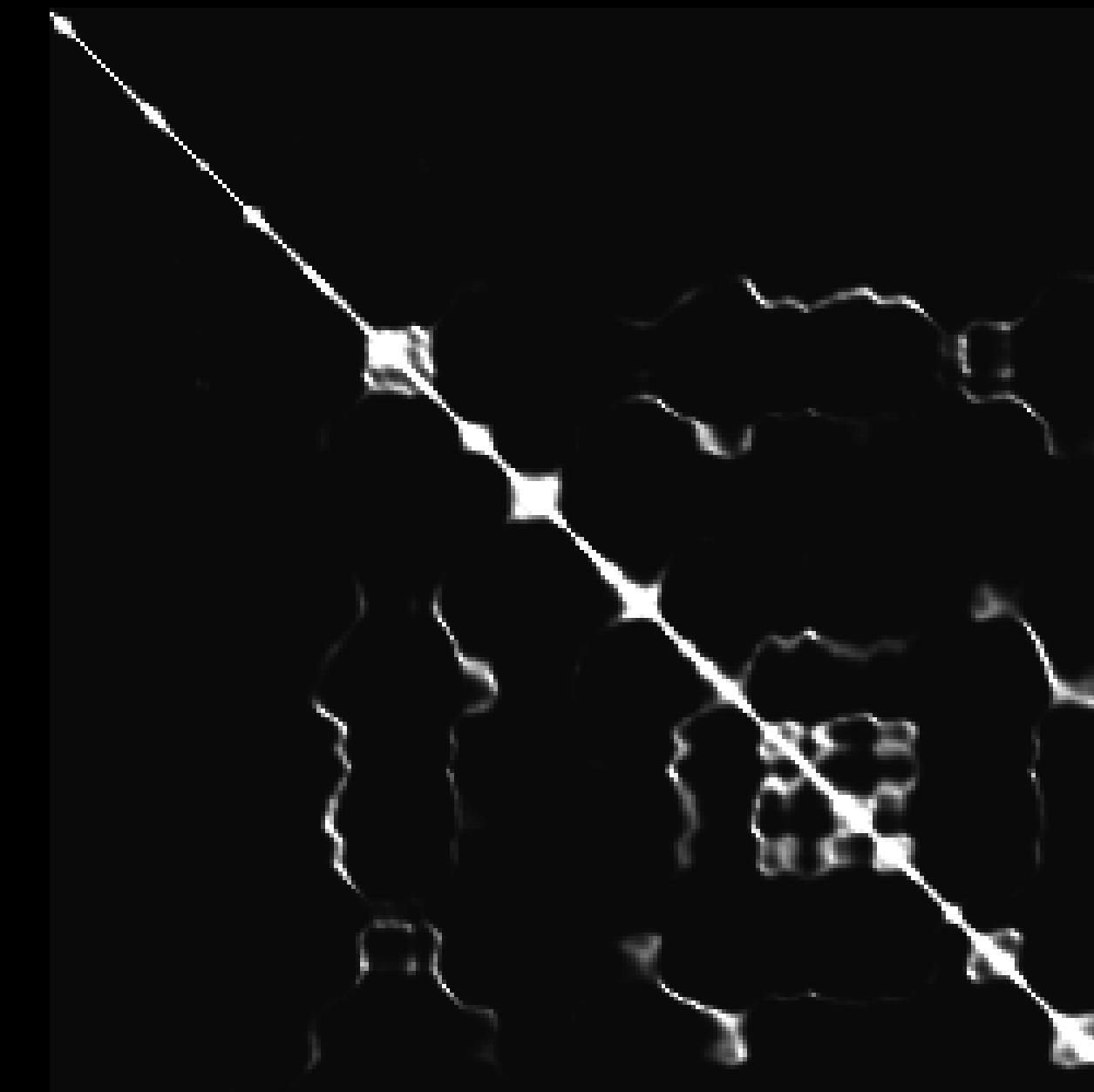
Finding good transitions

- Compute L_2 distance $D_{i,j}$ between all frames



Similar frames make good transitions

Markov chain representation

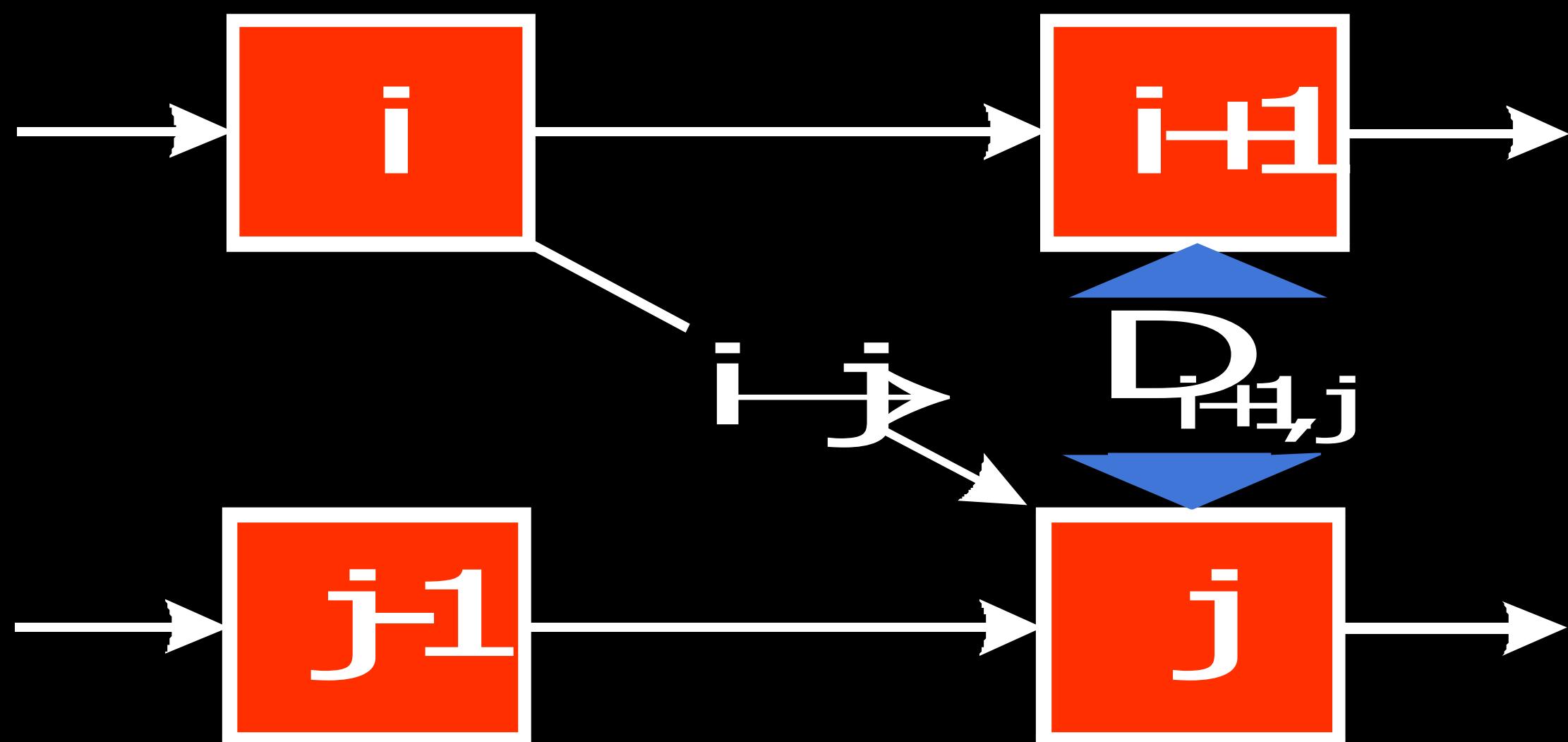


Similar frames make good transitions

Transition costs

- Transition from i to j if successor of i is similar to j
 - Cost function: $C_{i \rightarrow j} = D_{i+1, j}$

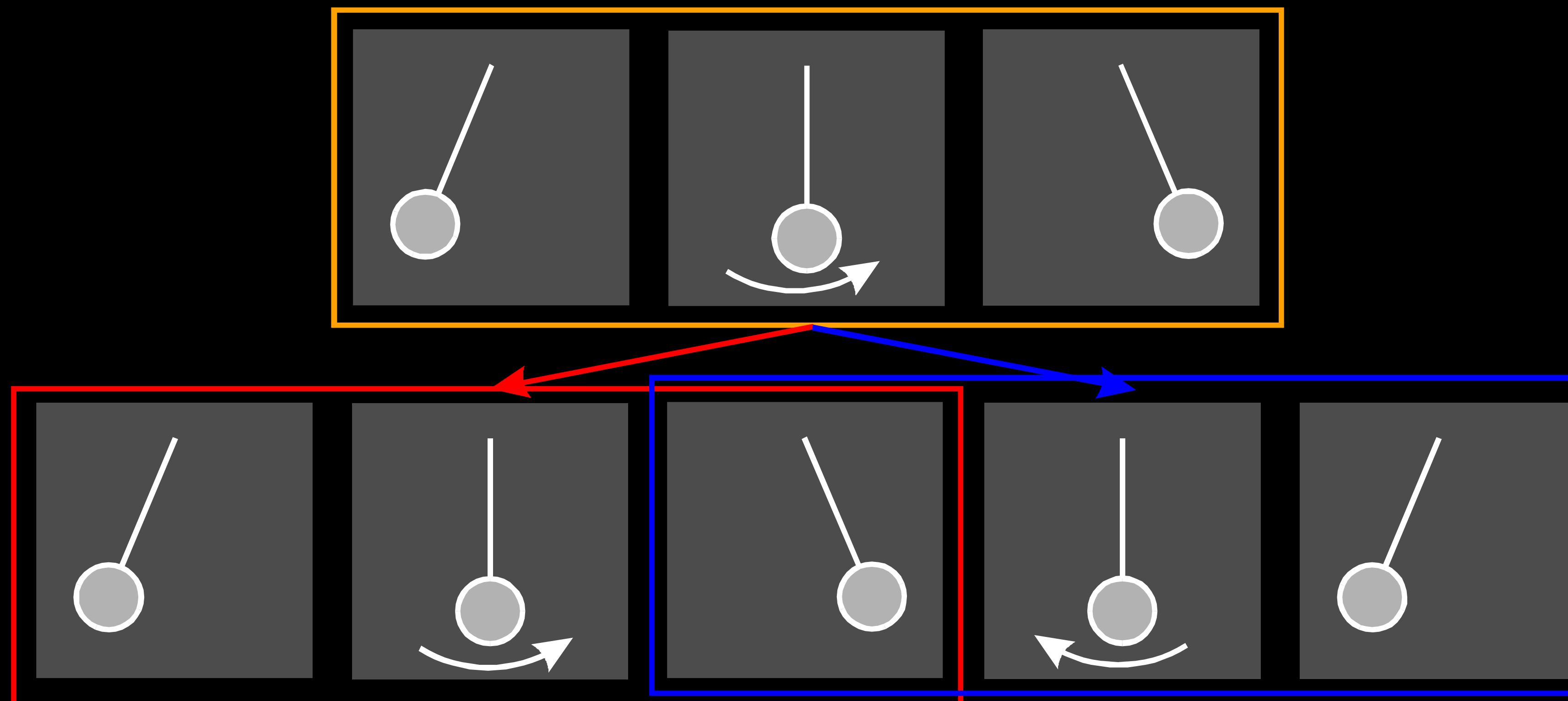
.



Preserving dynamics



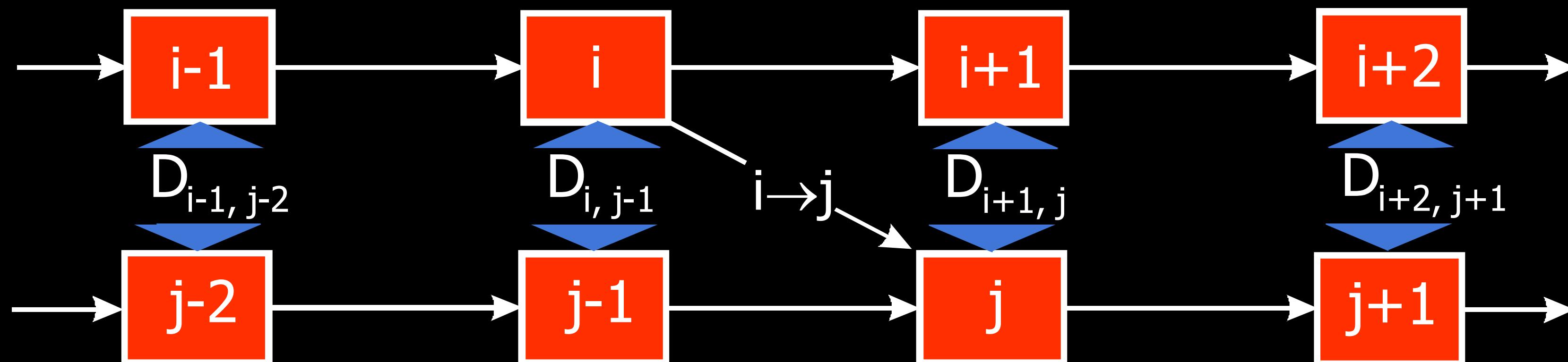
Preserving dynamics



Preserving dynamics

- Cost for transition $i \rightarrow j$

- $$C_{i \rightarrow j} = \sum_{k = -N}^{N-1} w_k D_{i+k+1, j+k}$$



Preserving dynamics – effect

- Cost for transition $i \rightarrow j$
 - $C_{i \rightarrow j} = \sum_{k = -N}^{N-1} w_k D_{i+k+1, j+k}$



Video portrait



- c.f. Harry Potter

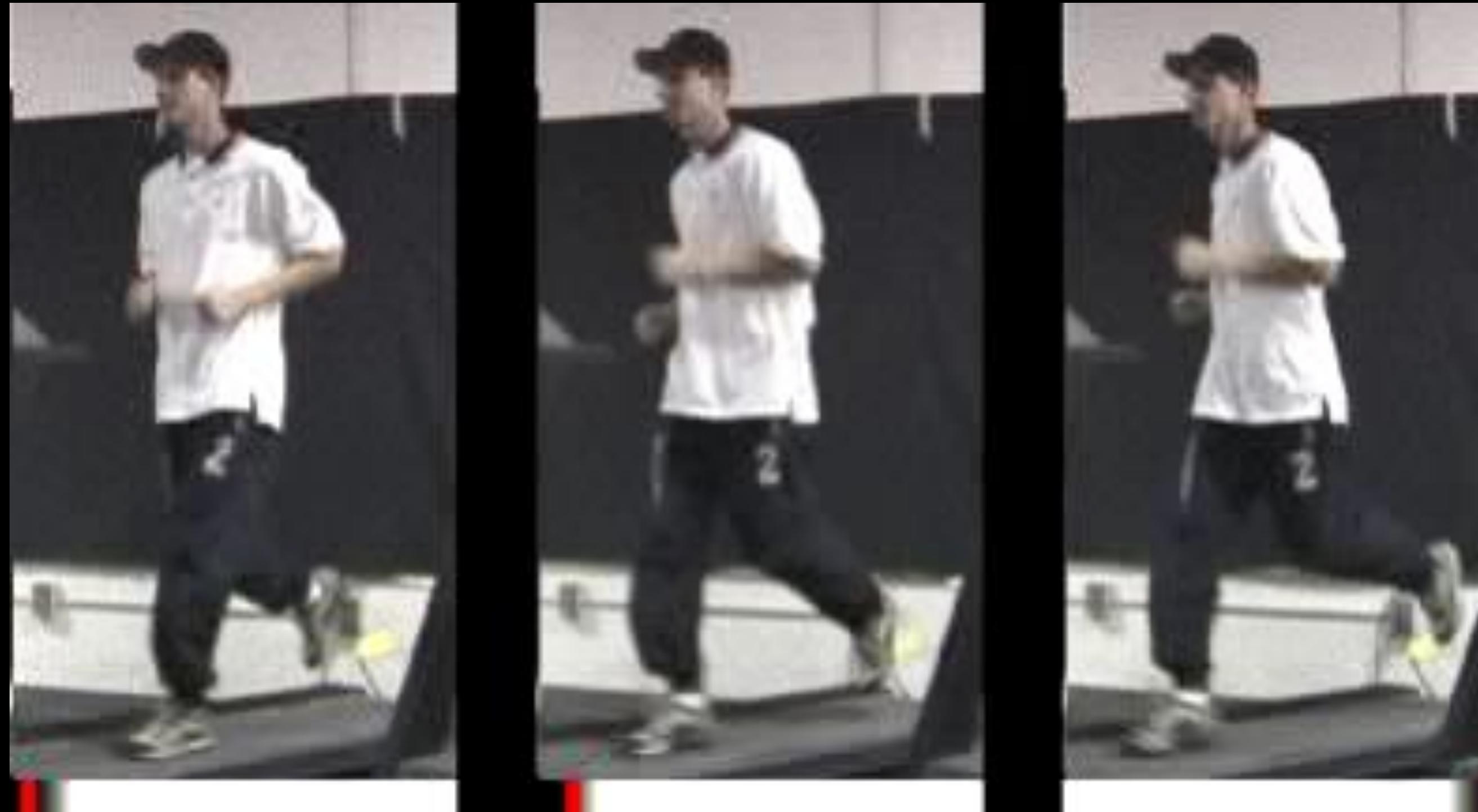
Region-based analysis

- Divide video up into regions



- Generate a video texture for each region

User-controlled video textures



slow

variable

fast

User selects target frame range

Video-based animation

- Like sprites
computer games
- Extract sprites
from real video
- Interactively control
desired motion



©1985 Nintendo of America Inc.



Video sprite extraction



blue screen matting
and velocity estimation



Video sprite control

- Augmented transition cost:

$$C_{i \rightarrow j}^{\text{Animation}} = \alpha C_{i \rightarrow j} + \beta \text{ angle}$$

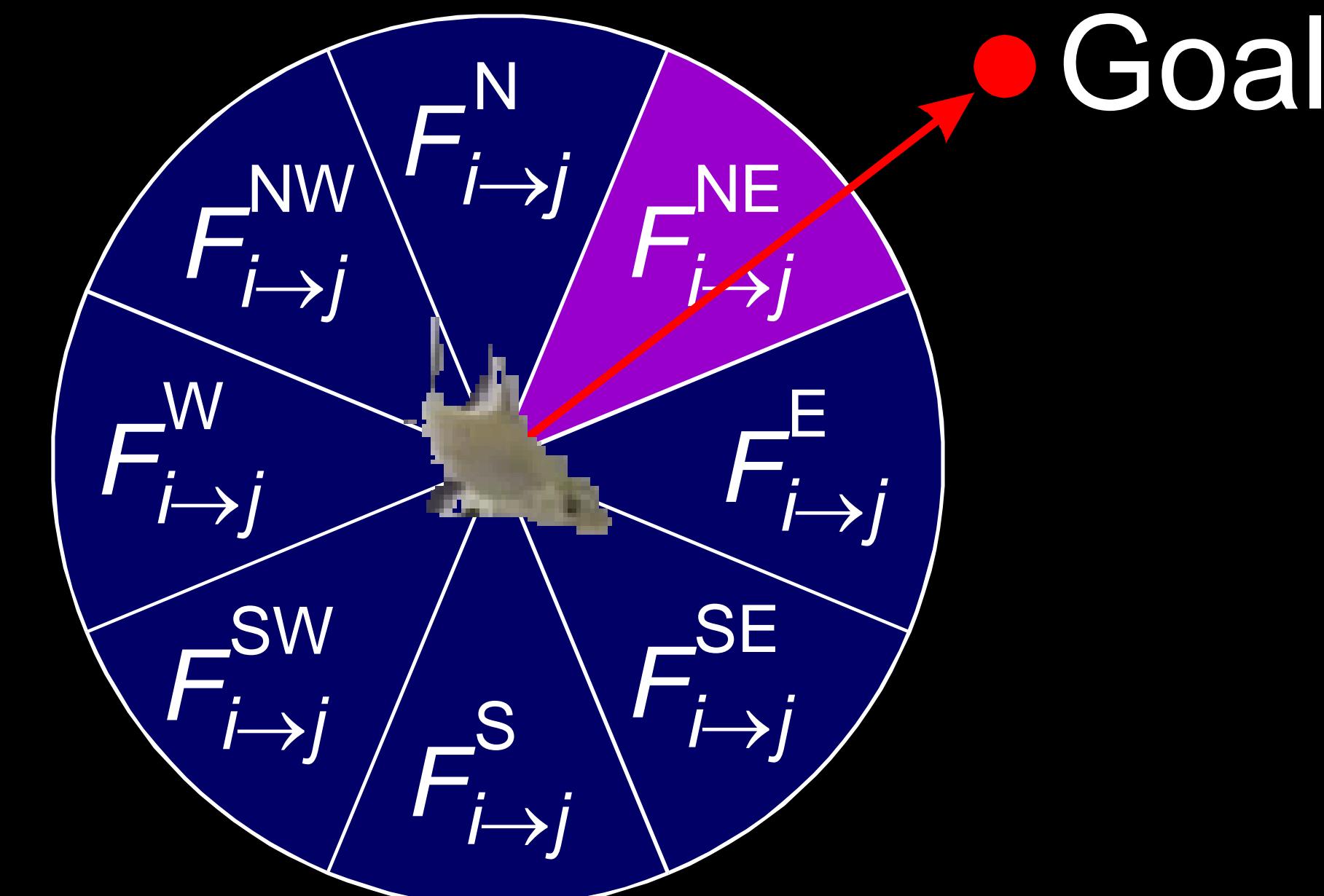
Similarity term Control term

vector to
mouse pointer

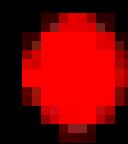
velocity vector

Video sprite control

- Need future cost computation
- Precompute future costs for a few angles.
- Switch between precomputed angles according to user input
- [GIT-GVU-00-11]



Interactive fish



Summary / Discussion

- Some things are relatively easy

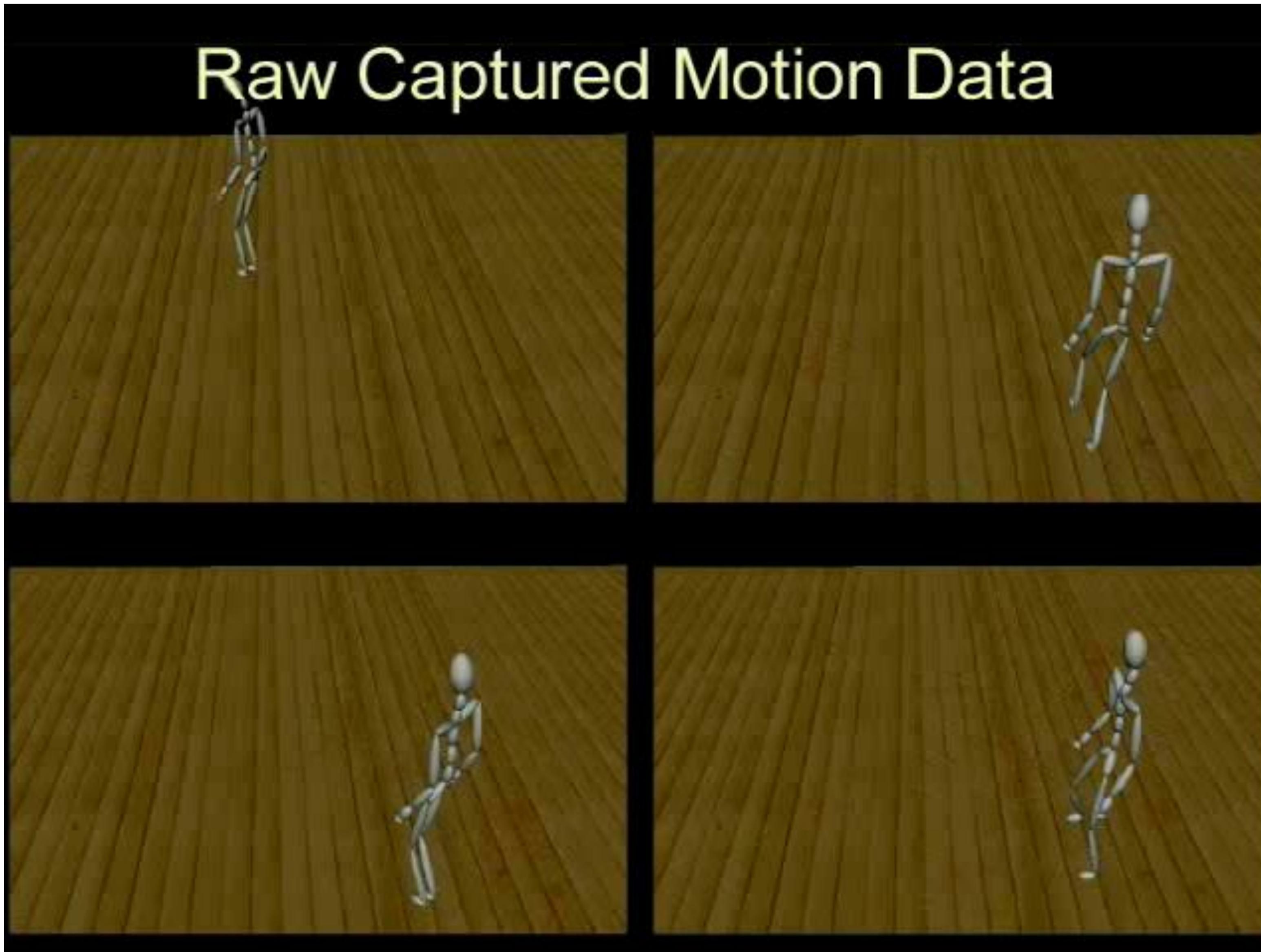


Discussion

- Some are hard



Motion Caption Database

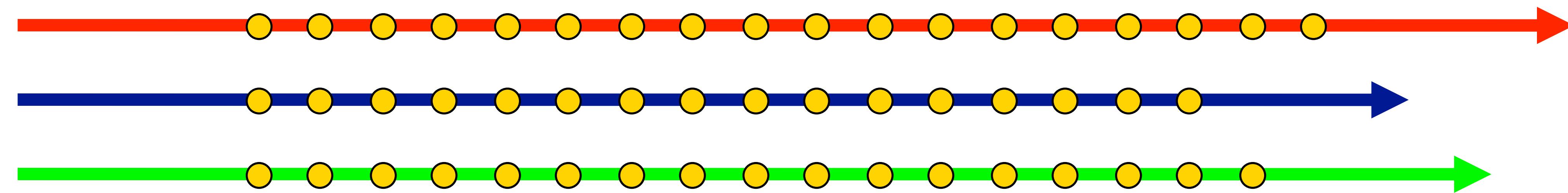


Sketch Interface

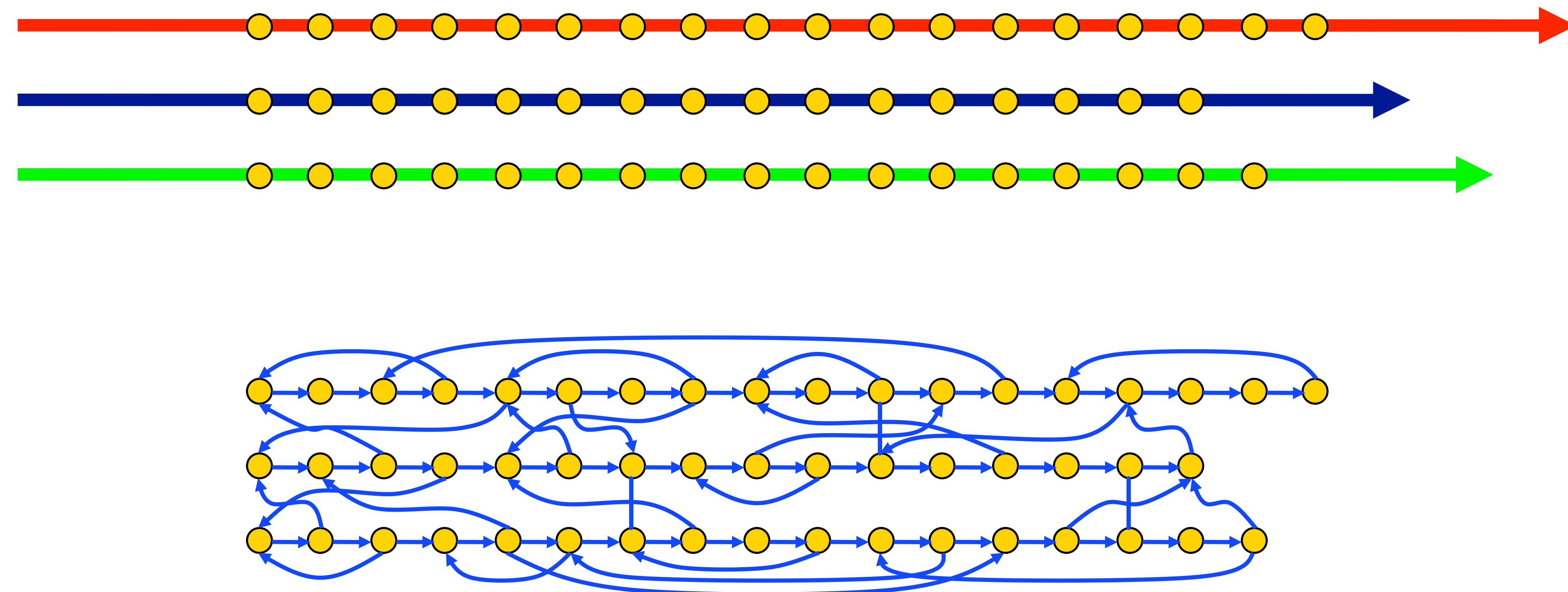


Jehee Lee, Jinxiang Chai, Paul Reitsma, Nancy Pollard, Jessica Hodgins SIGGRAPH 2002

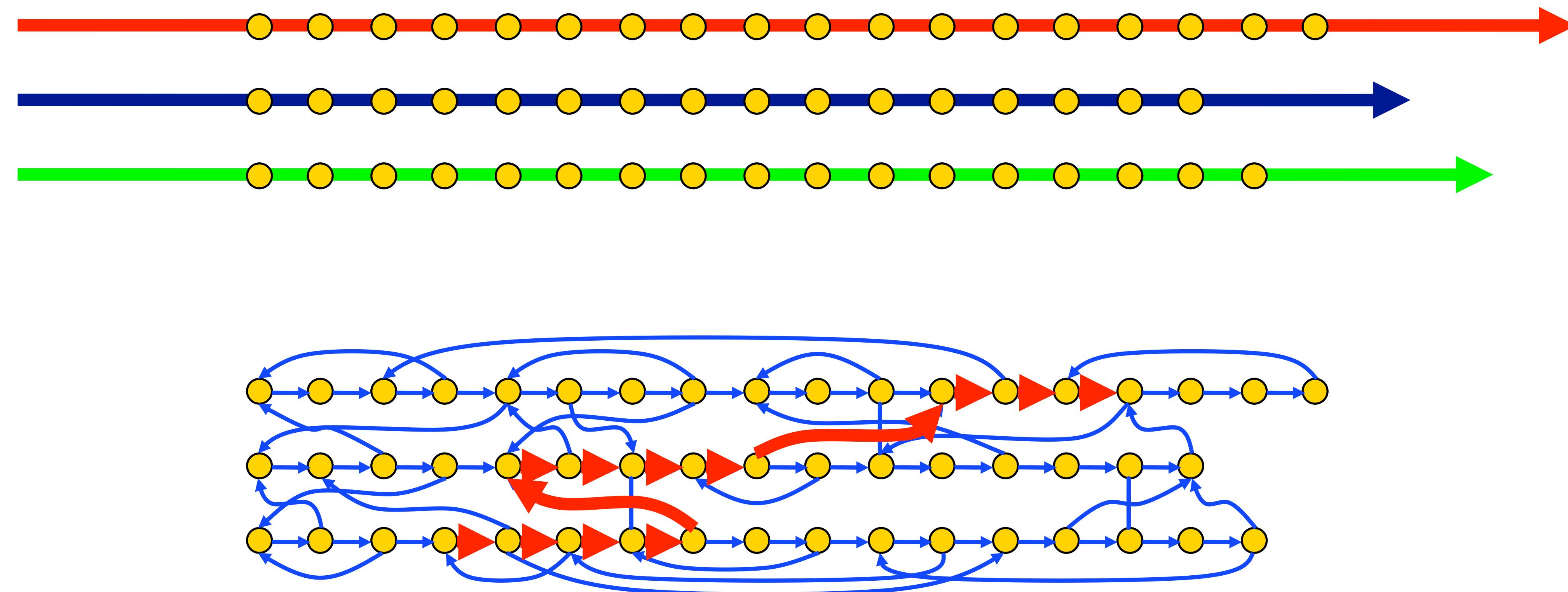
Unstructured Input Data



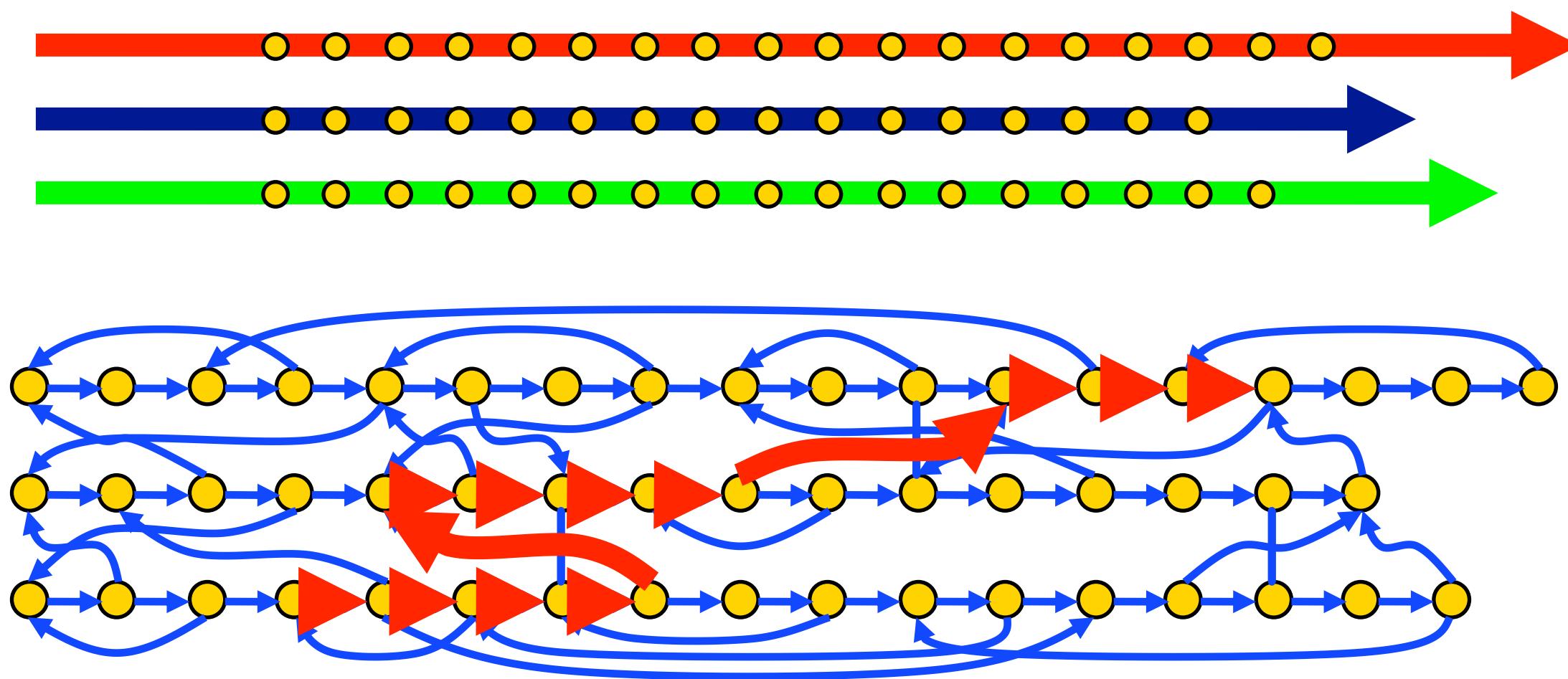
Connecting Transitions



Search to Find Path



Motion Data for Rough Terrain



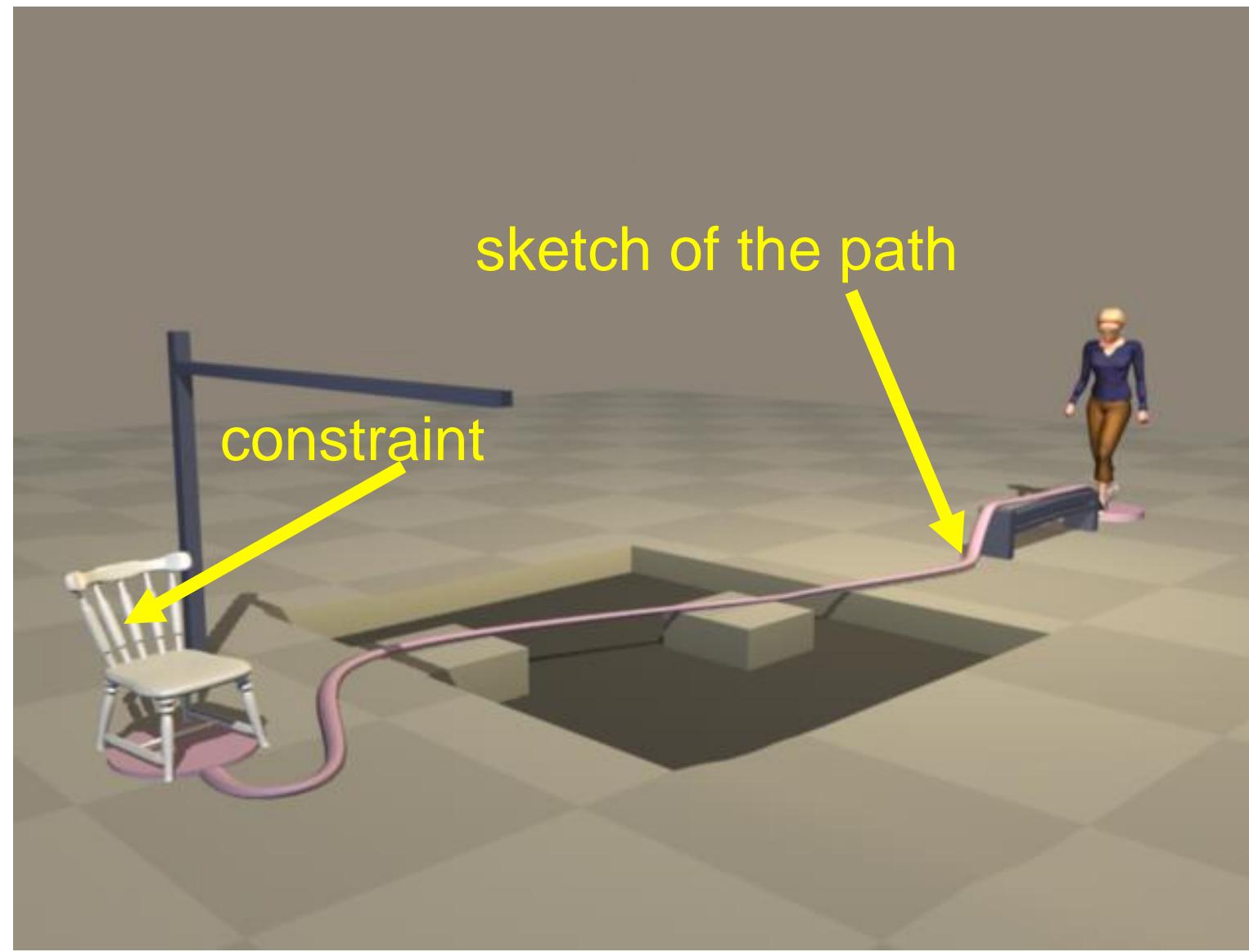
What did they learn?

Databases do not need to be nearly as big as we would have guessed
~5 minutes is sufficient for walking

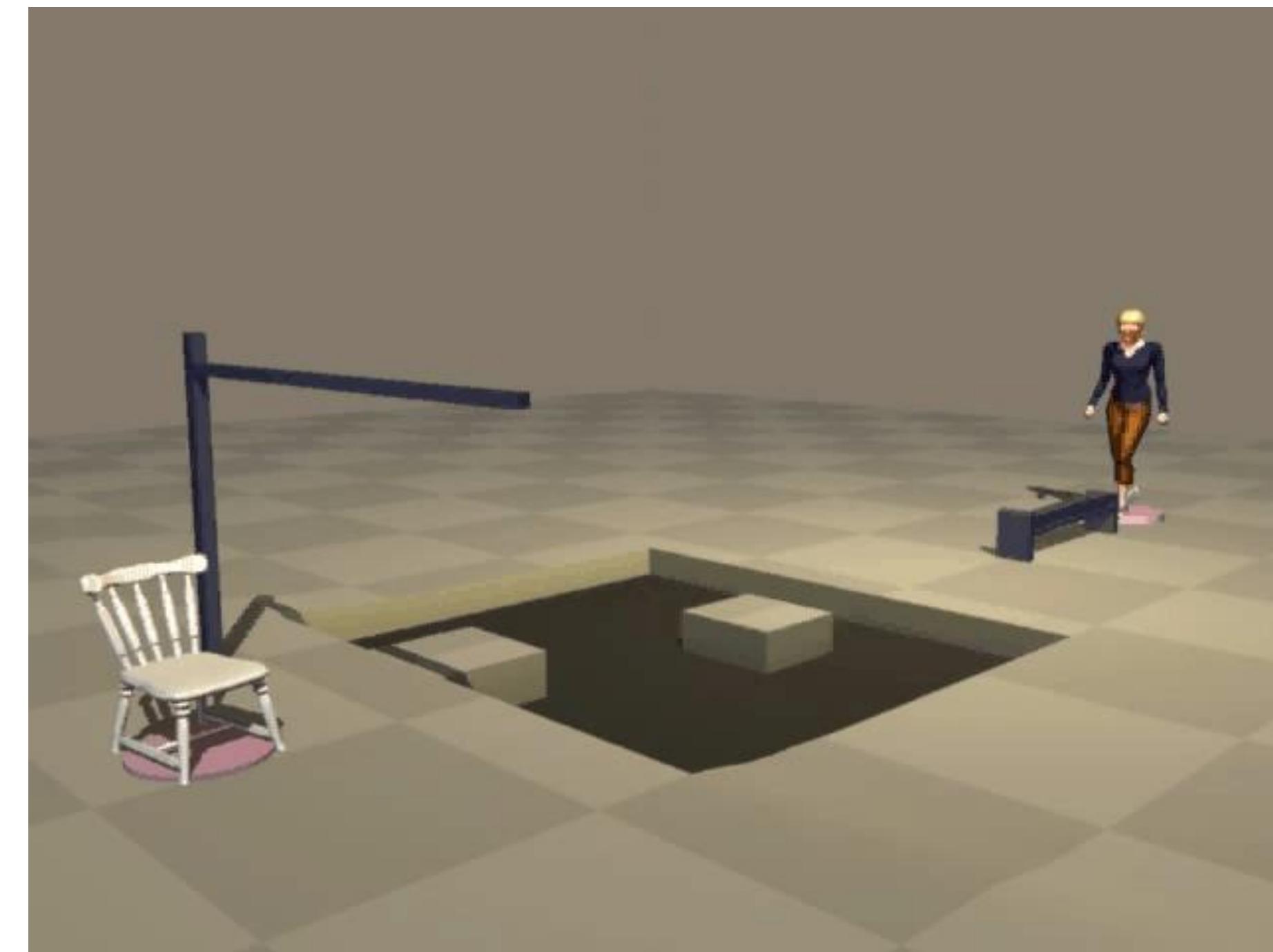
Planning in the space of possible behaviors (as represented by the data) is quite efficient

Interpolated Motion Graphs

Sketch of the path

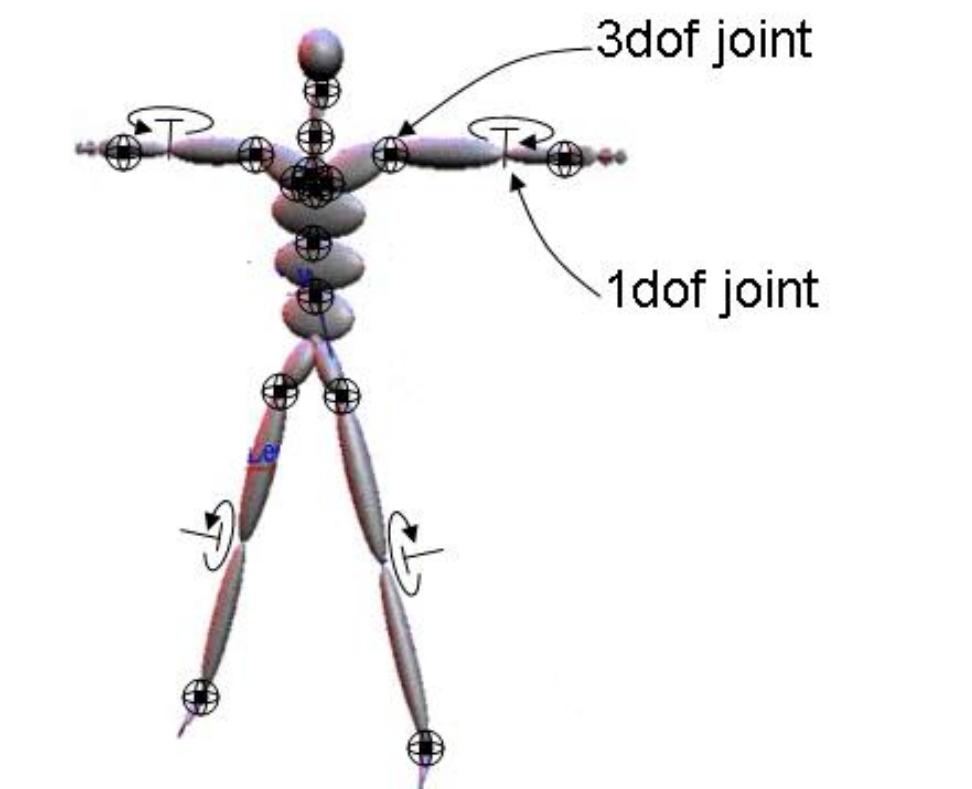
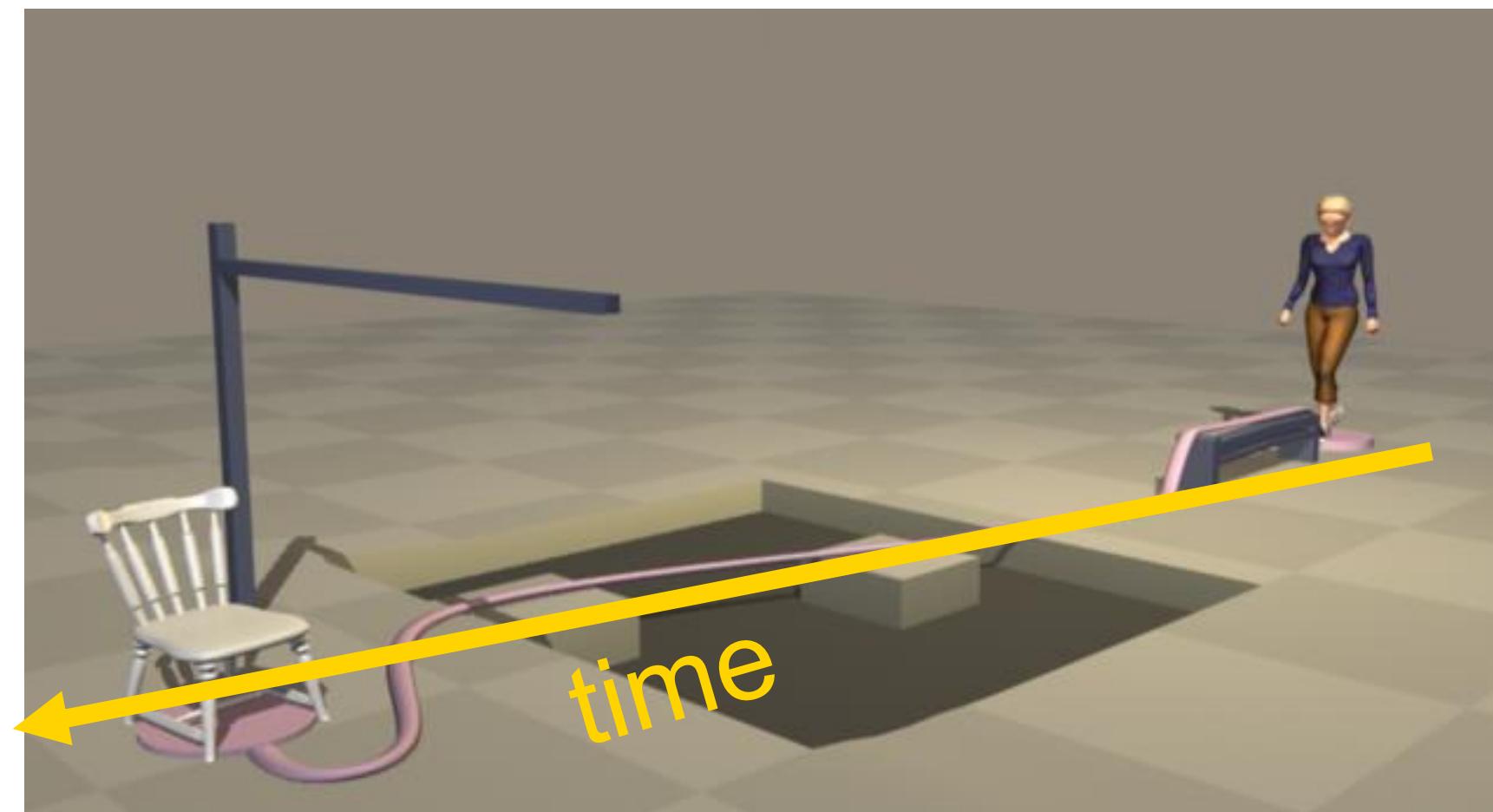


Synthesized motion



Optimization Problem

Unknowns: poses of the character over time

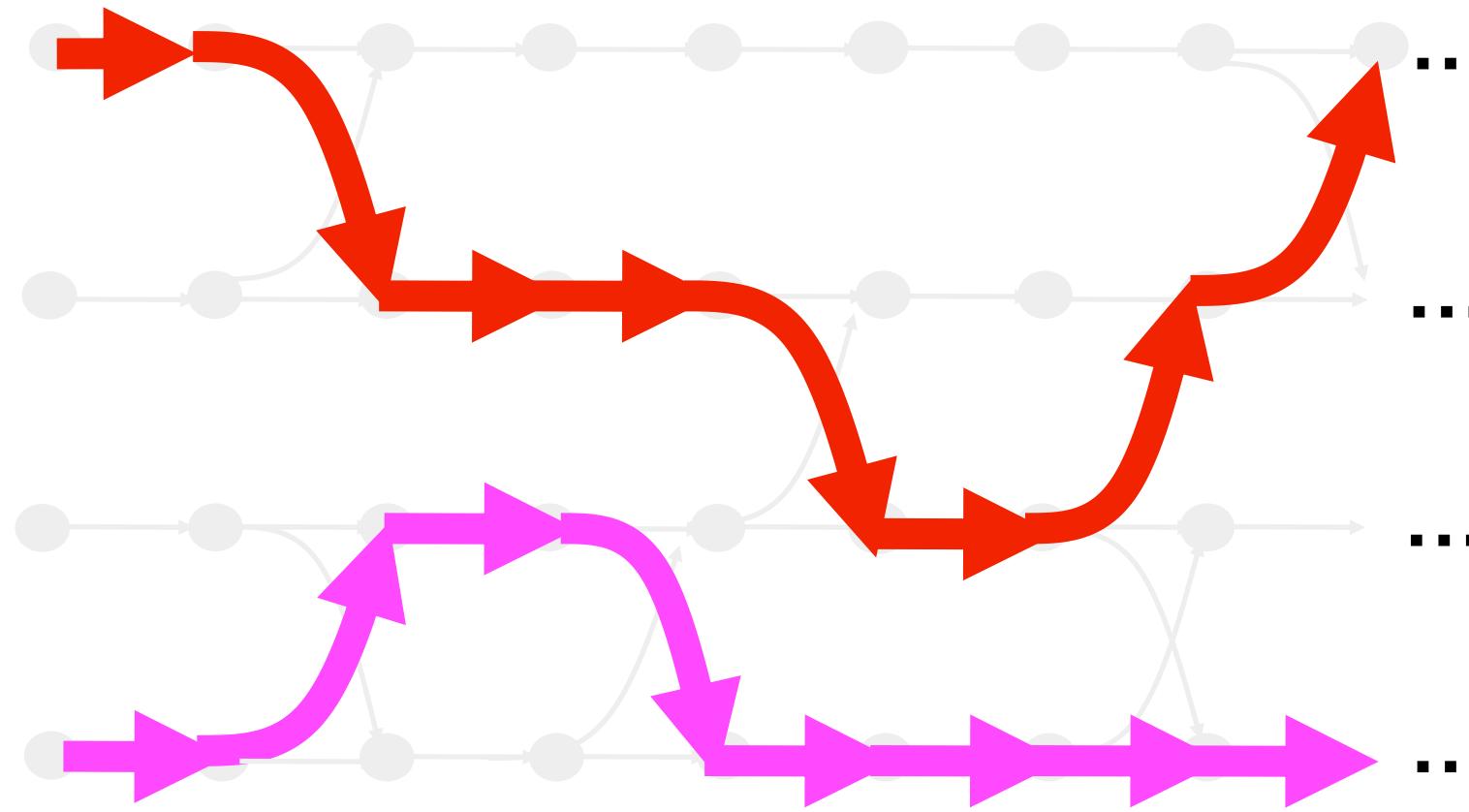


Minimize: sum of squared torques

Constraints: user, environmental and physics

Naïve Search space is very large: $50 \times T$ unknowns

Key Idea 1: Interpolated Motion Graphs

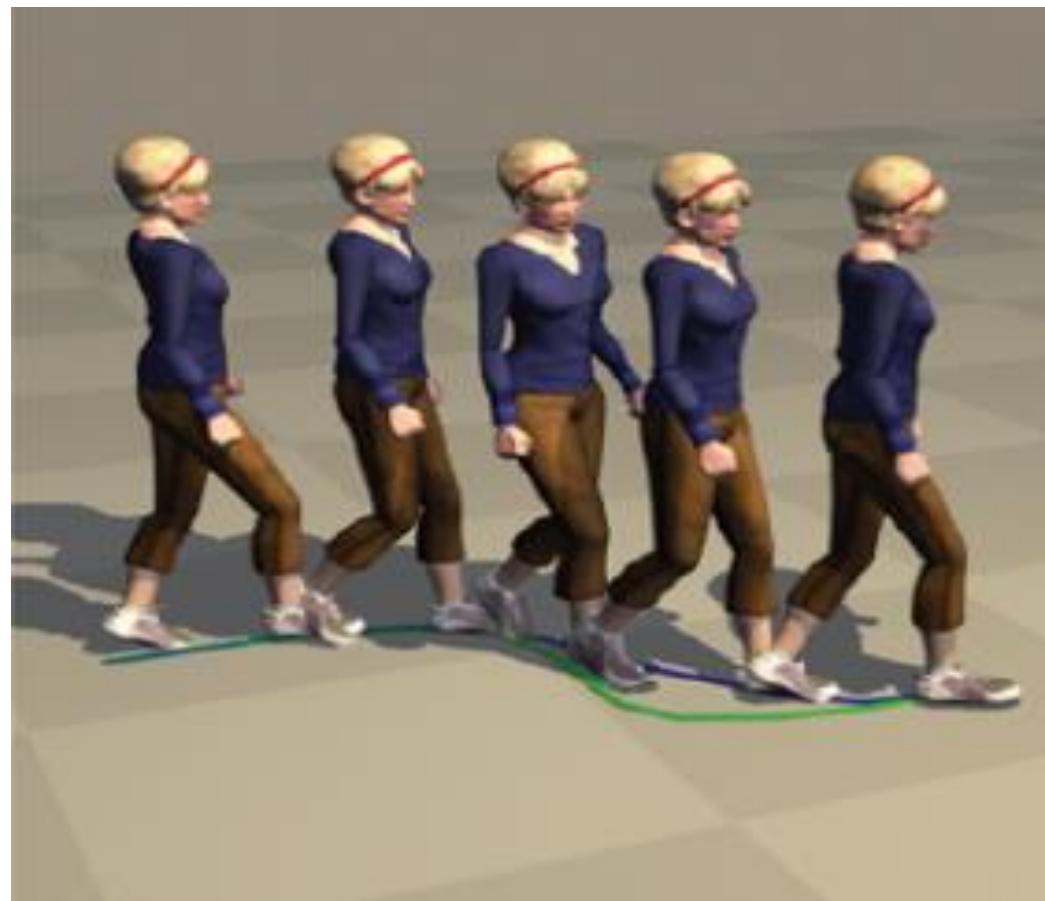


Compact representation of motion

Long, multi-behavior motions

Novel motions

Interpolation



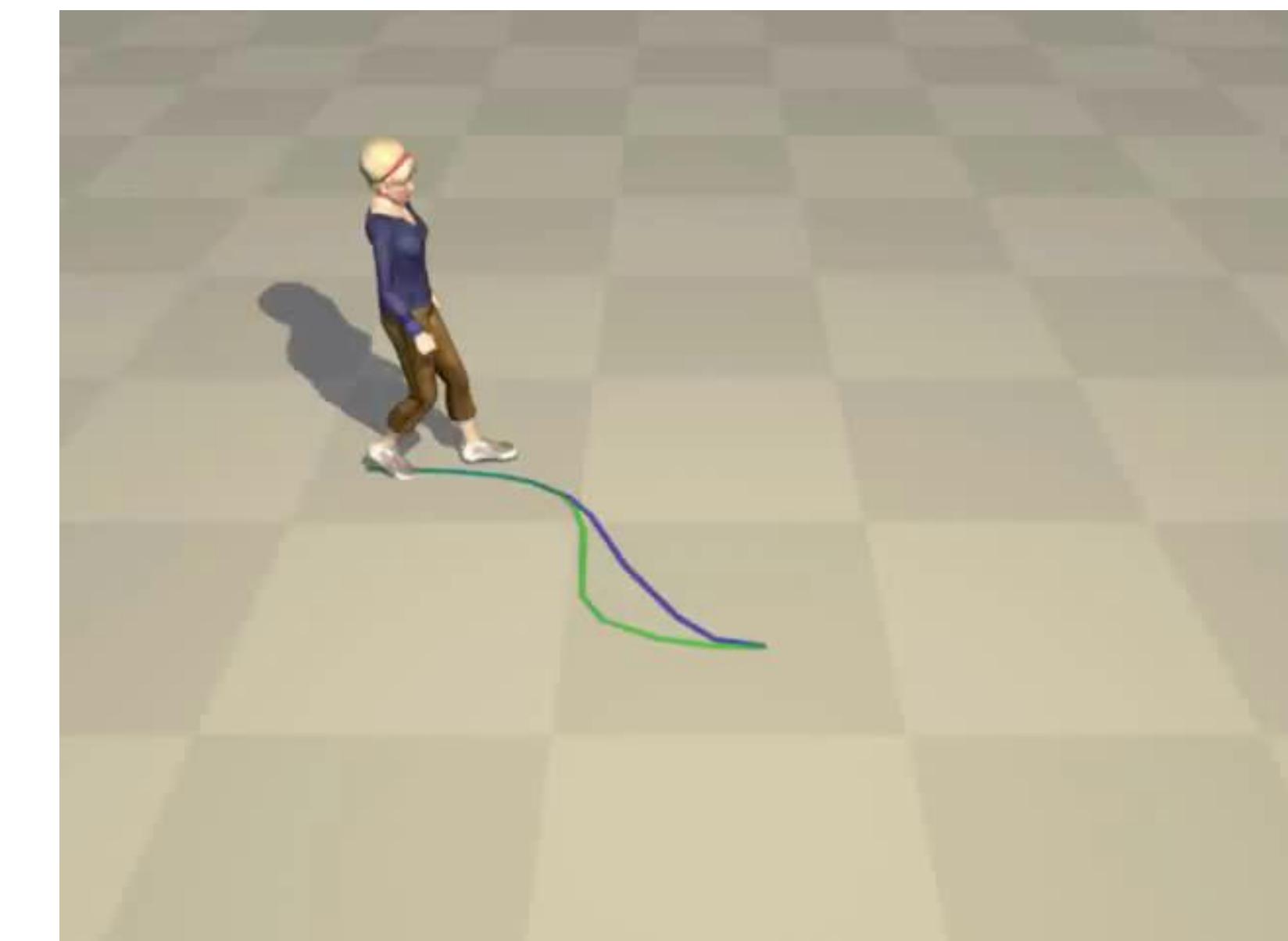
no interpolation



with interpolation



no interpolation



with interpolation

Optimality

**sub-optimal
solutions**



optimal solution



Alla Safonova & Jessica Hodgins,
SIGGRAPH 2007

Texture

- Texture depicts spatially repeating patterns
- Many natural phenomena are textures



radishes



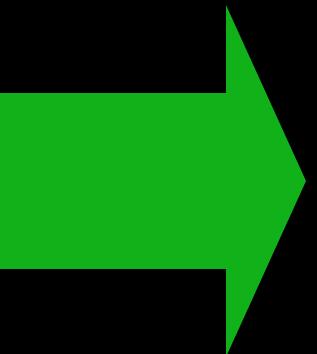
rocks



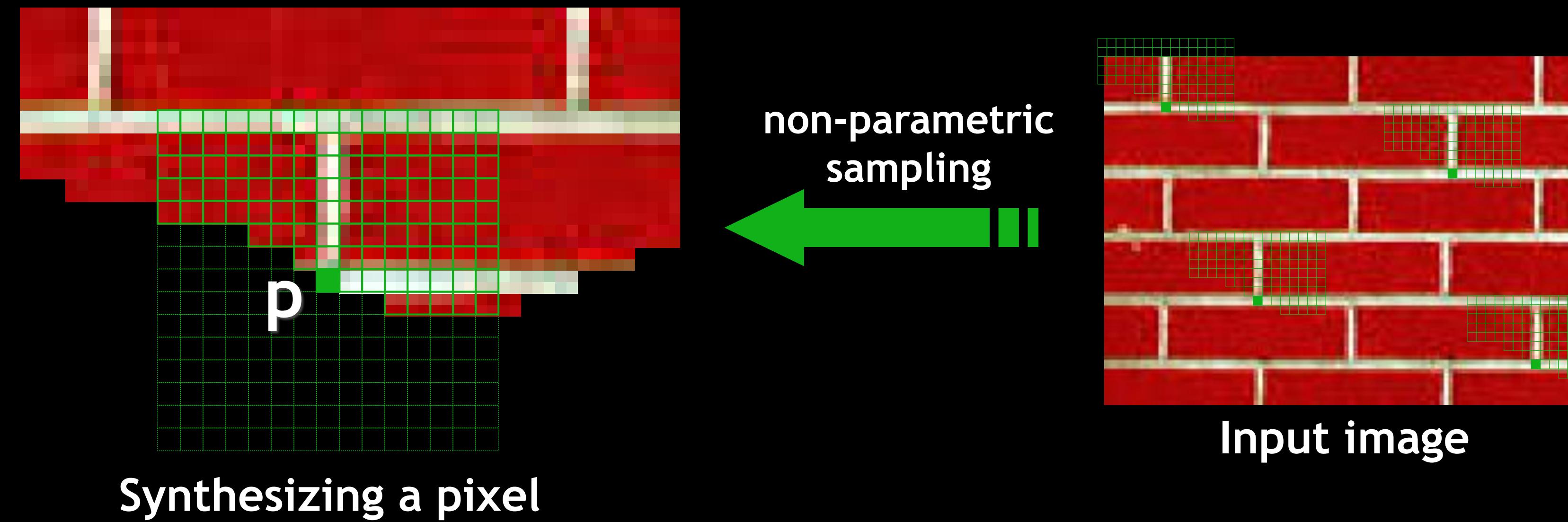
yogurt

Texture Synthesis

- Goal of Texture Synthesis: create new samples of a given texture
- Many applications: virtual environments, hole-filling, texturing surfaces



Efros & Leung Algorithm (ICCV 1999)

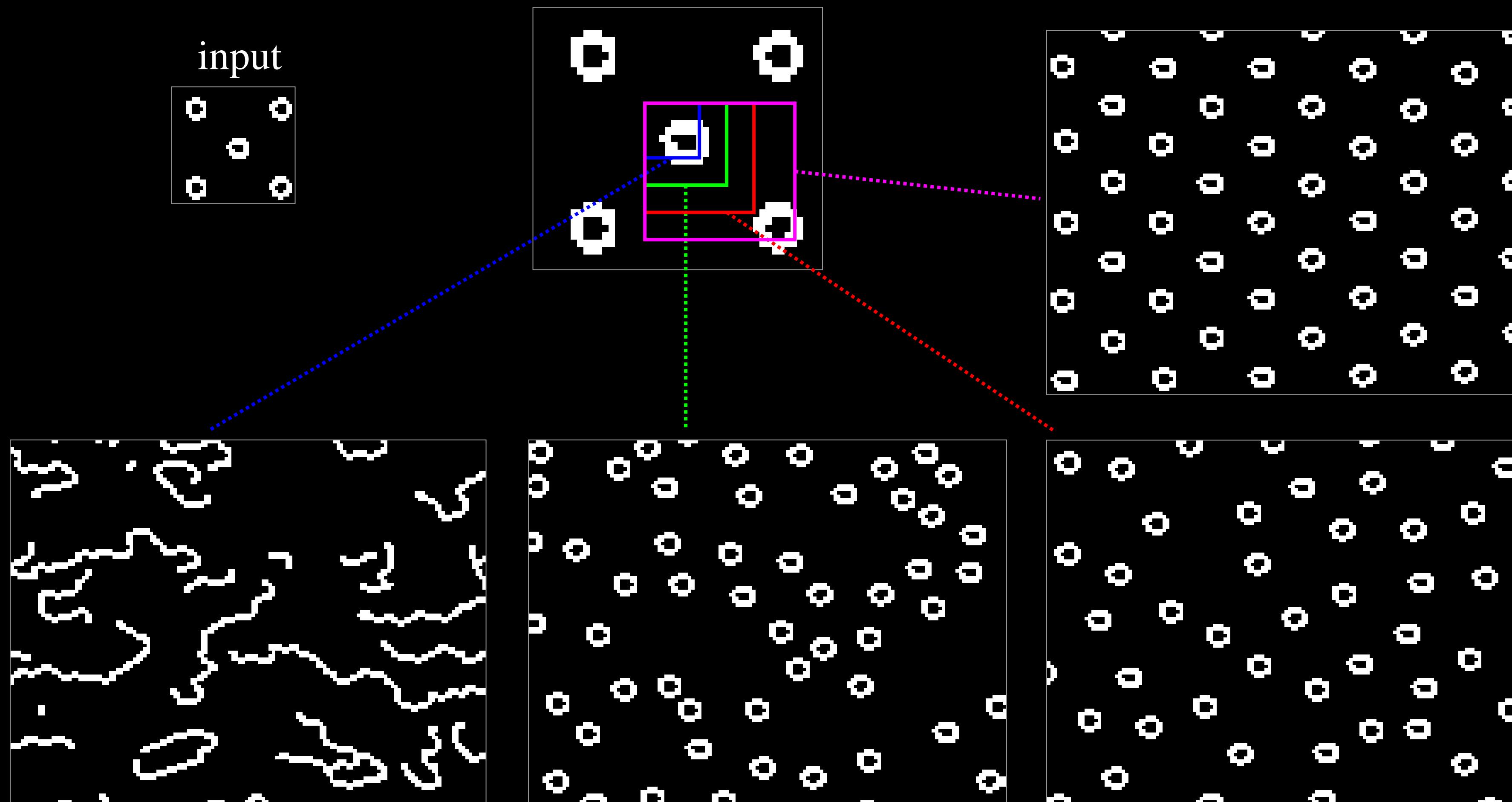


- Assuming Markov property, compute $P(p|N(p))$
 - Building explicit probability tables infeasible
 - Instead, we *search the input image* for all similar neighborhoods — that's our pdf for p
 - To sample from this pdf, just pick one match at random

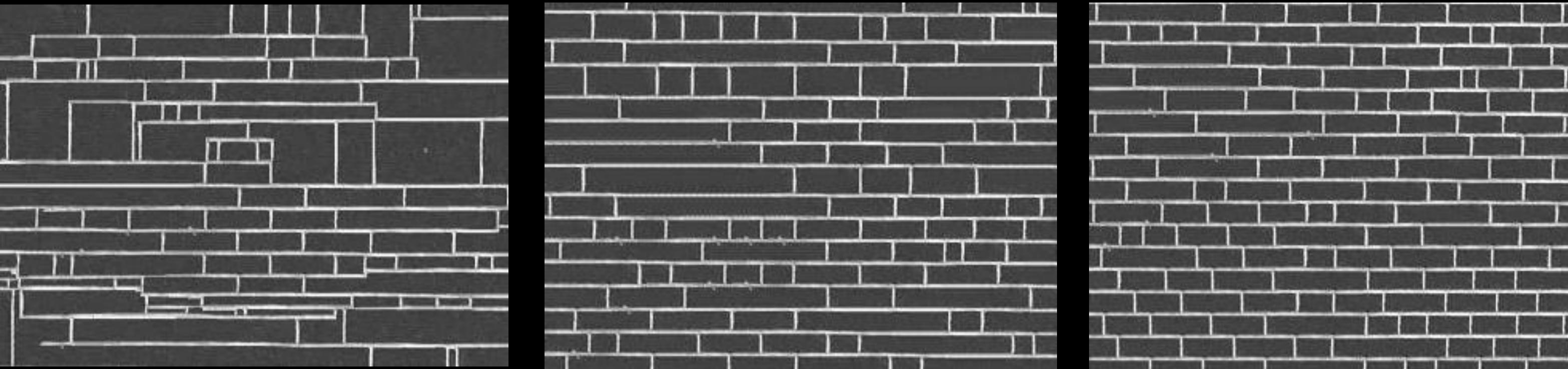
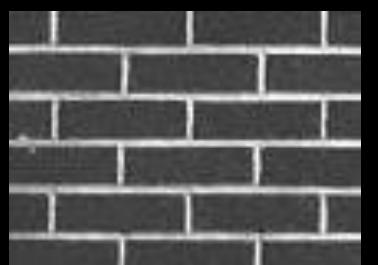
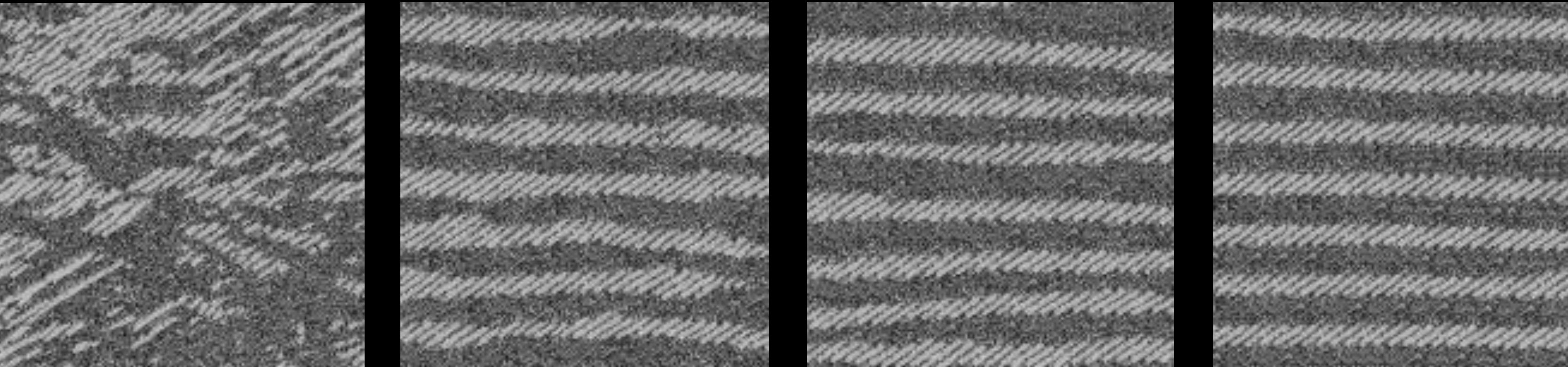
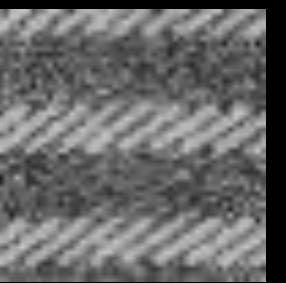
Some Details

- Growing is in “onion skin” order
 - Within each “layer”, pixels with most neighbors are synthesized first
 - If no close match can be found, the pixel is not synthesized until the end
- Using *Gaussian-weighted* SSD is very important
 - to make sure the new pixel agrees with its closest neighbors
 - Approximates reduction to a smaller neighborhood window if data is too sparse

Neighborhood Window



Varying Window Size

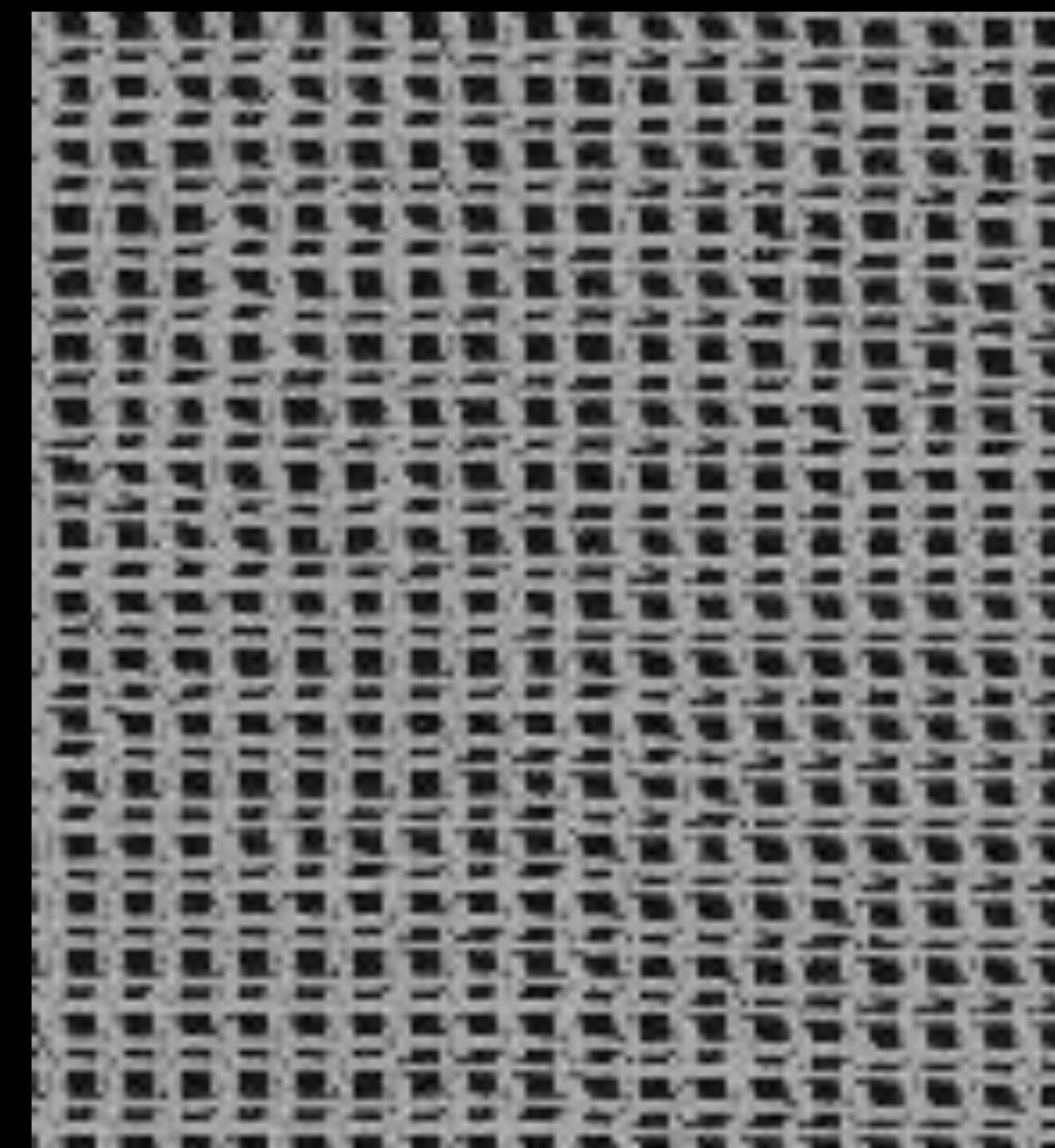
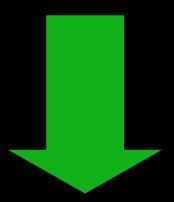
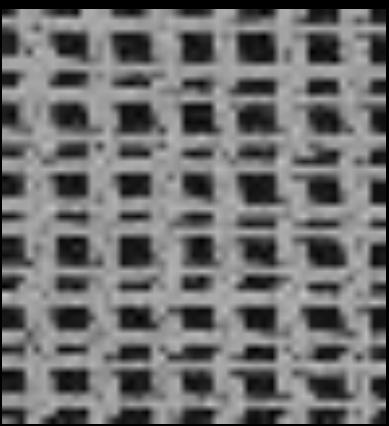


Increasing window size

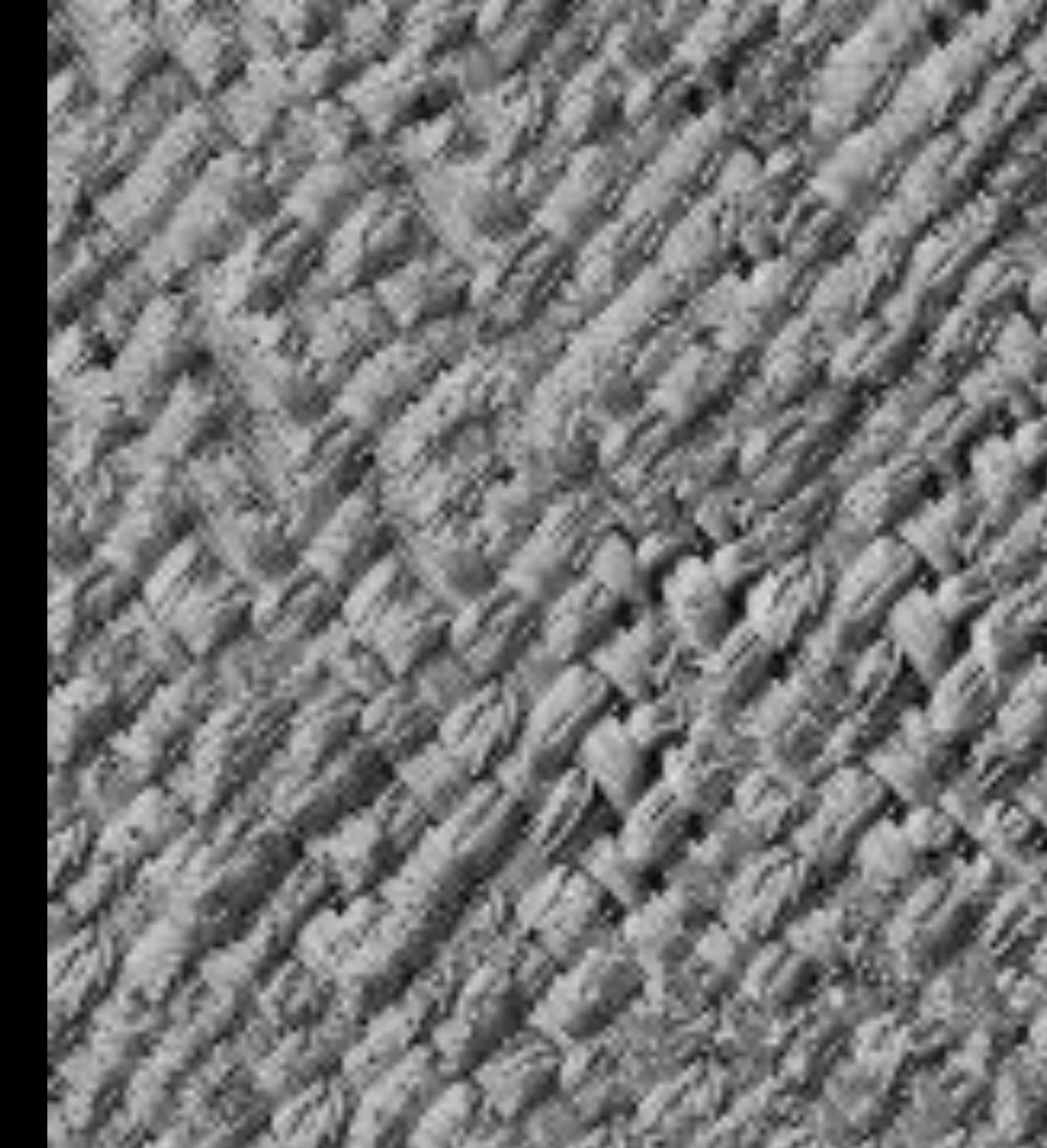
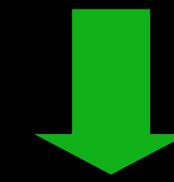
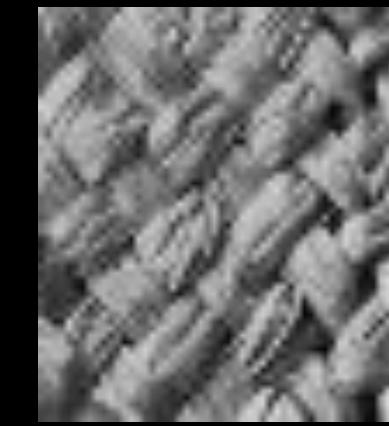


Synthesis Results

french canvas

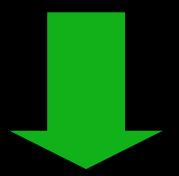


rafia weave

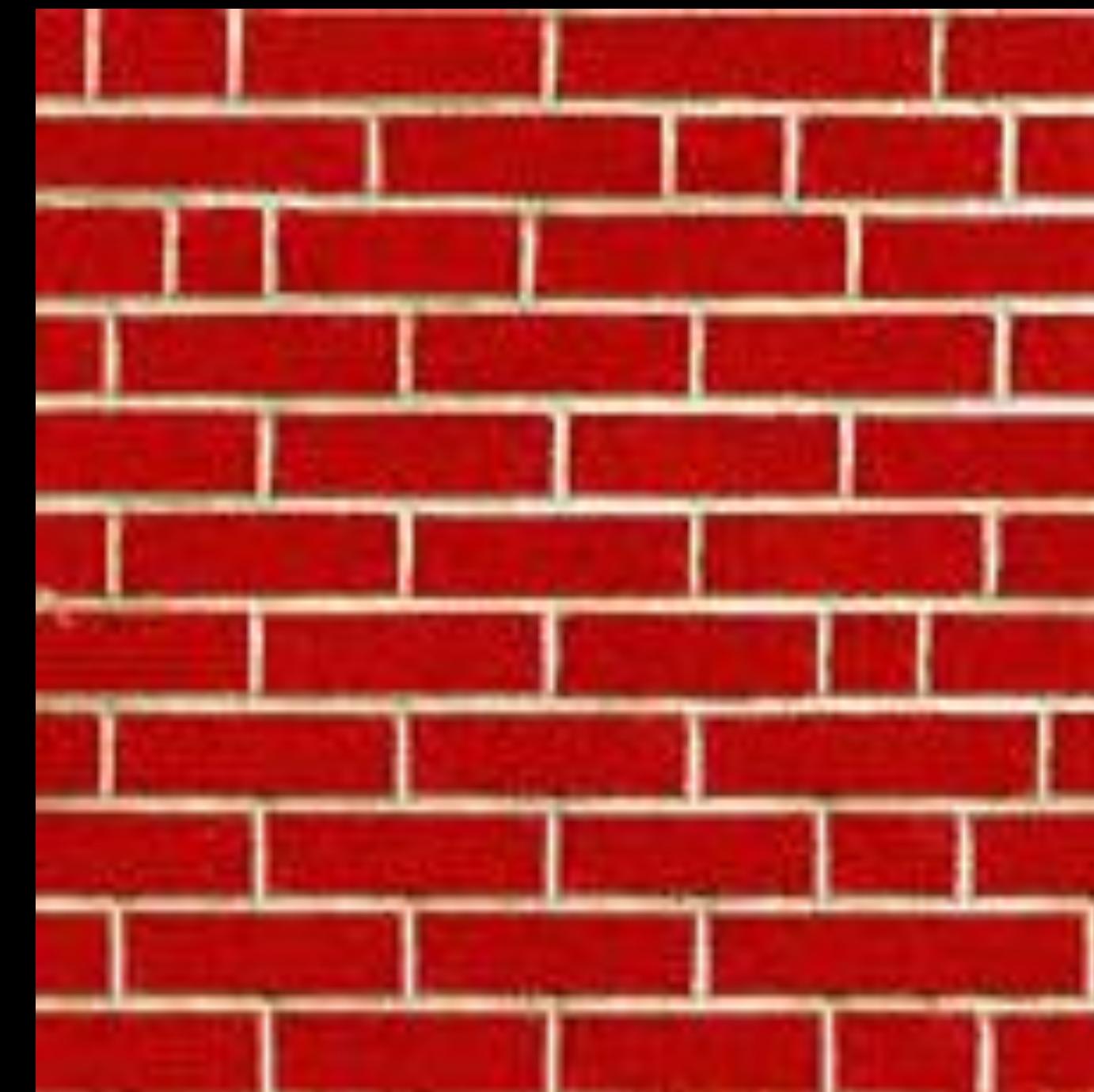
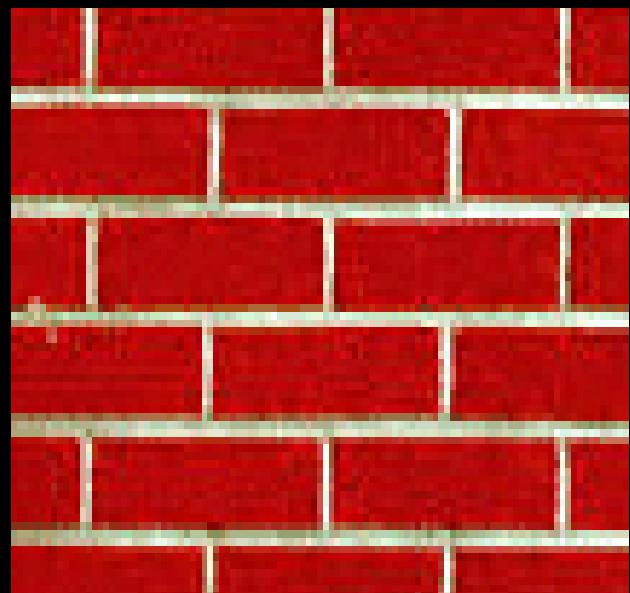


More Results

white bread



brick wall



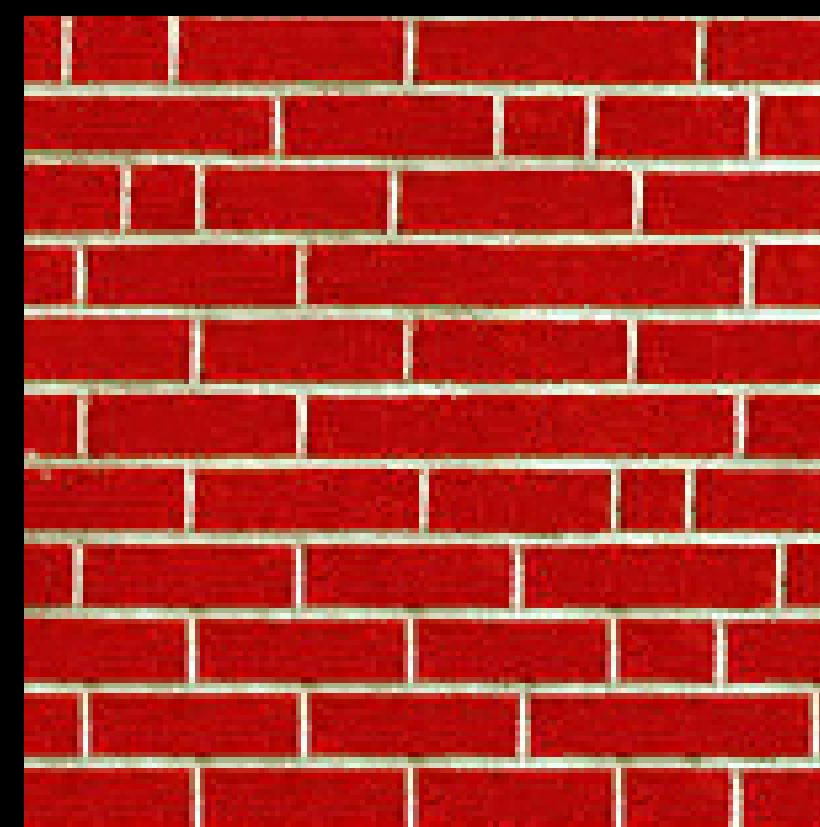
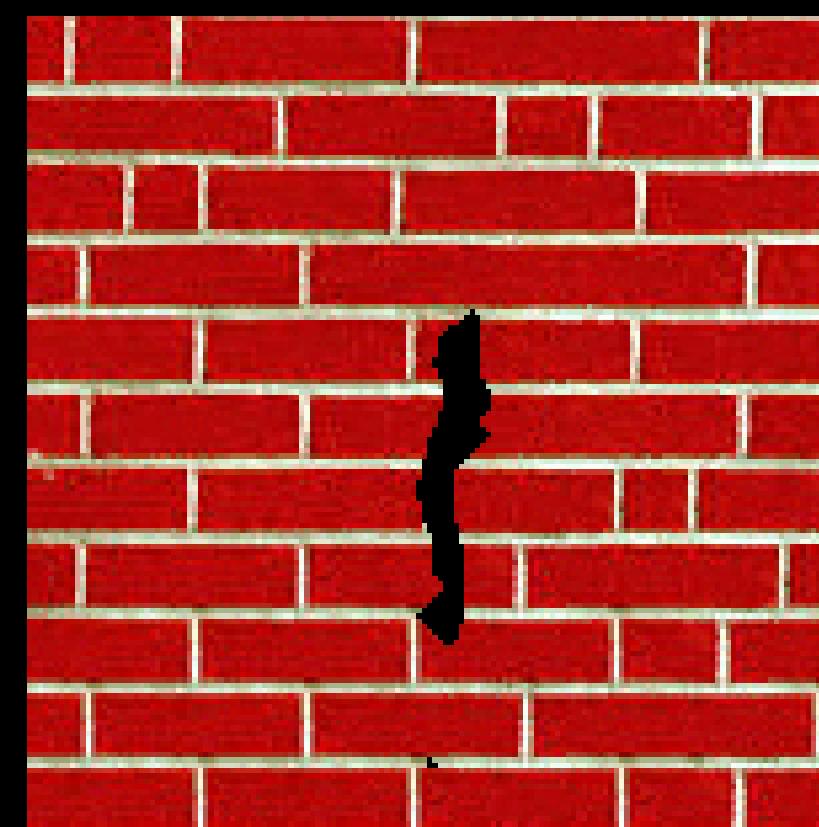
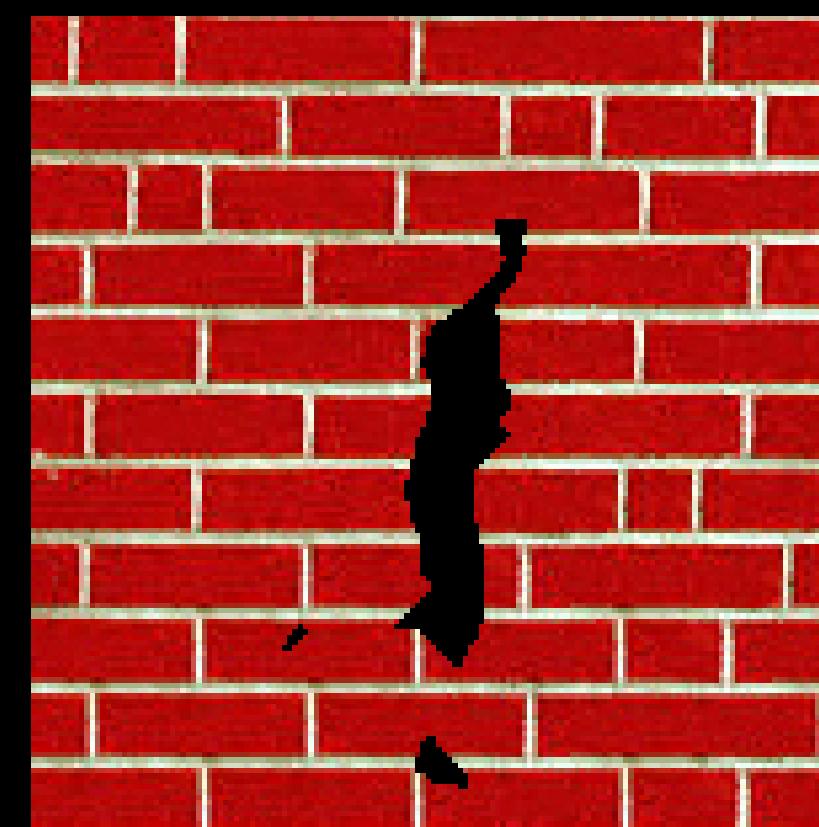
Homage to Shannon

oring in the unsensational. Mr. Dick Gephardt was fair enough to offer a painful riff on the looming Senate's decision. "I only asked, "What's your opinion?" A heartfelt sigh followed. "It's a story about the emergence of a new set of values against Clinton. "Boy, did he know how to bring people about continuing to support him." Gephardt began, patiently observing that the legal system had been "shattered" with this latest tangle.

"I b'vopobodr'lehA'wC'... ar'... e
eagle per's otot... la'tini' hht... lmr' thunne...
...as "he'd c'... er' eior' feori' A'...
h'... ry' te f' fit' a'... tif' foear' f'...
t'ae' t' diab'... h'... h'... j'... m'... te opm'...
te fit' seu'... v'... bnt' u'... rn' hrist' d'... h'... pnr'...
n' id' r'f' p'... e'... g'... d'... g'... l'... g'... l'... ir'... l'... enr'...
utonuc'... f'... h'... s'... il'...
"jthenly'... n'... A'... f'... er'... eloaeunnh'... n'... n'...
'dthf' p'... di'... l'... in'... e'... o'... ha' muuny'... r'... b'... a'...
or'... f'... fa'... if'... J'... h'... ooegahmtt'... sy'...
t'... s'... s'... h'... sk'... as'... k'... y'... s'... j'... n'... C'... u'... e'... t'... f'... n'...
e'... i'... s'... t'... h'... k'... y'... s'... A'... h'... n'... f'... r'...
t'... in'... t'... r'... f'... a'... e'... c'... d'... t'... e'... l'... r'... C'... a'... r'... s'...
s'... C'... n'... h'... r'... t'... e'... l'... y'... a'... p'... n'... s'... A'... h'... n'... f'... r'...
s'... t'... h'... i'... n'... e'... g'... i'... l'... i'... j'... i'... ur'... a'... b'... e'... n'... n'... f'... Q'... n'...
t'... h'... o'... t'... e'... t'... y'... e'... n'... C'... i'... r'... y'... i'... l'... i'... j'... i'... ur'... a'...
t'... dy'... l'... f'... e'... r'... o'... e'... r'... e'... n'... r'... t'... h'... f'... f'... b'... l'... i'... r'... r'... n'... "...'... b'... t'...
t'... e'... [ps'... the]... t'... g'... e'... r'... i'... o'... f'... H'... s'... v'... p'... n'... r'... p'... d'...
t'... n'... A'... a'... r'... e'... r'... o'... j'... o'... m'... t'... n'... f'... r'... c'... n'... t'... t'... o'... t'... h'... p'... n'... s'... t'...
t'... a'... m'... r'... i'... r'... o'... p'... r'... e'... r'... t'... e'... l'... i'... t'... u'... r'... i'...
t'... g'... m'... r'... i'... r'... e'... r'... t'... e'... l'... i'... t'... u'... r'... i'...
t'... B'... u'... n'... j'... u'... n'... o'... r'... m'... l'... e'... p'... s'... a'... h'... b'... +'... C'... o'... a'... t'... i'... a'... w'... w'... r'...
t'... e'... n'... a'... t'... n'... e'... n'... h'... h'... m'... s'... f'... a'... n'...

l'... thaim'... them'... "Whnephartfe hartfelintomimen'...
fel ck Clirtioout omaim thartfelinsfaut' s anentd'...
the ry onst wartfe lck Gephtoomimeationl sigab'...
Chiooufit Clinut Cll riff on, hat's yordn, parut tly'...
ons yoontonsteht wasked, paim t sahe loo' riff on I'...
nskoneplcourtfeas leil A' nst Clii, "Wleontongal s'...
k Cirtioouirtfepe ong pme abegal fartfenstemem'...
tiensteneltrydt telemephinsverdt was agemen'...
ff ons artientont Cling peme asurtfe atish, "Boui'...
nal s fartfelt sig pedrlrdt ske abounutie aboutoo'...
tfeaonewas vous abovonthardt thatins fain, ped, '...
ains, them, pabout wasy arfuit couitly d, l n A h'...
ble emthringbooreme agas fa bontinsyst Clinut'...
ory about continst Clipeopinst Cloke agatiff out O'...
stome minemen tly ardt beorabou n, thenly as t G'...
cons faimeme Diontont wat coutlyohgans as fan'...
ien, phrtfaul, "Wbaut cout congagal cominingai'...
mifmst Cliry abon al coountha emungairt tf ouni'...
Whe looorystan loontieph, intly on, theoplegatick'...
mul tatiesontly atie Diontiomt wal s f thegæ ener'...
mthahæst's enenhñmas fan, "intchthorw abons w'...

Hole Filling



Extrapolation

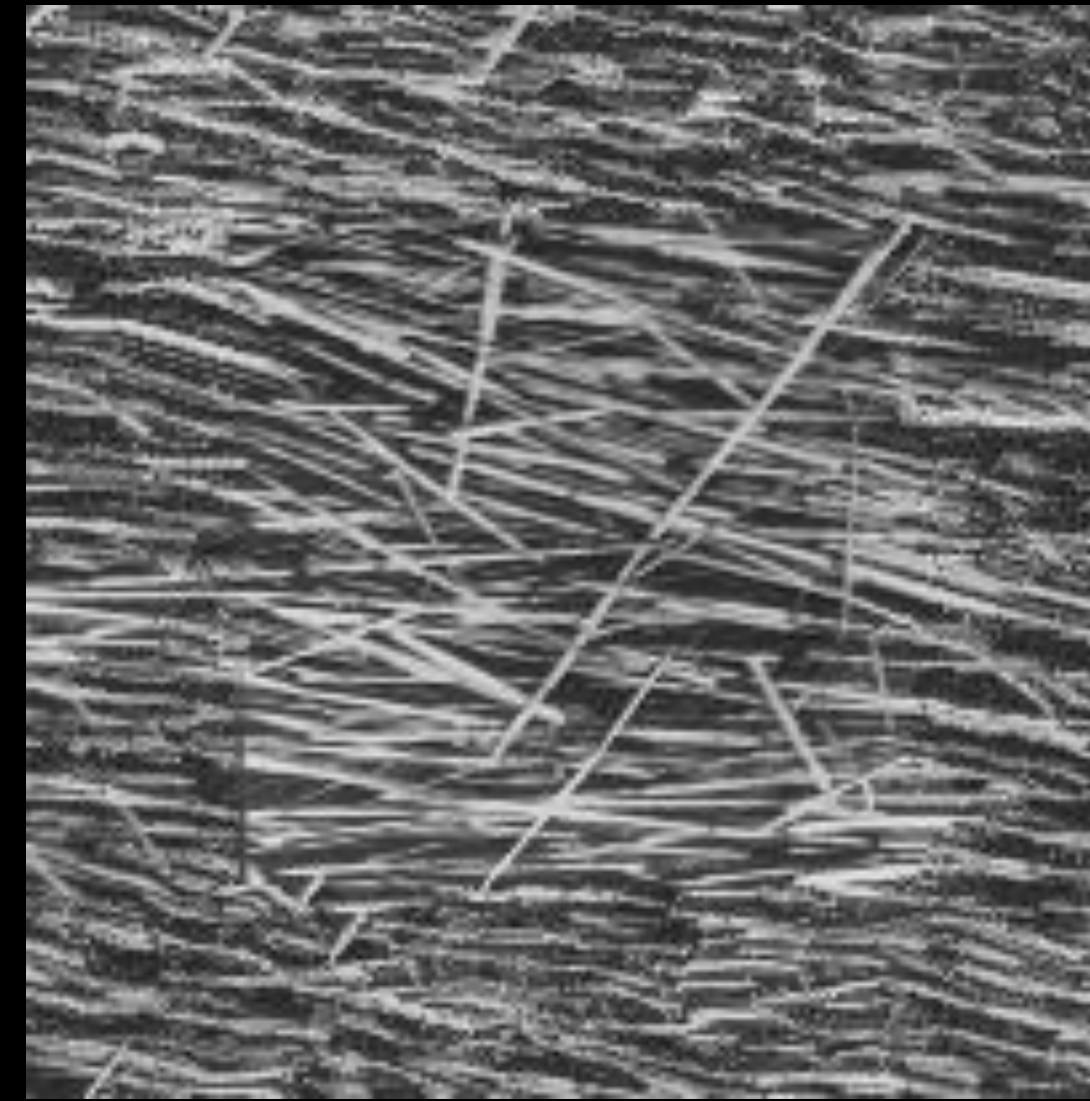


Image Analogies

Aaron Hertzmann^{1,2}

Chuck Jacobs²

Nuria Oliver²

Brian Curless³

David Salesin^{2,3}

¹New York University

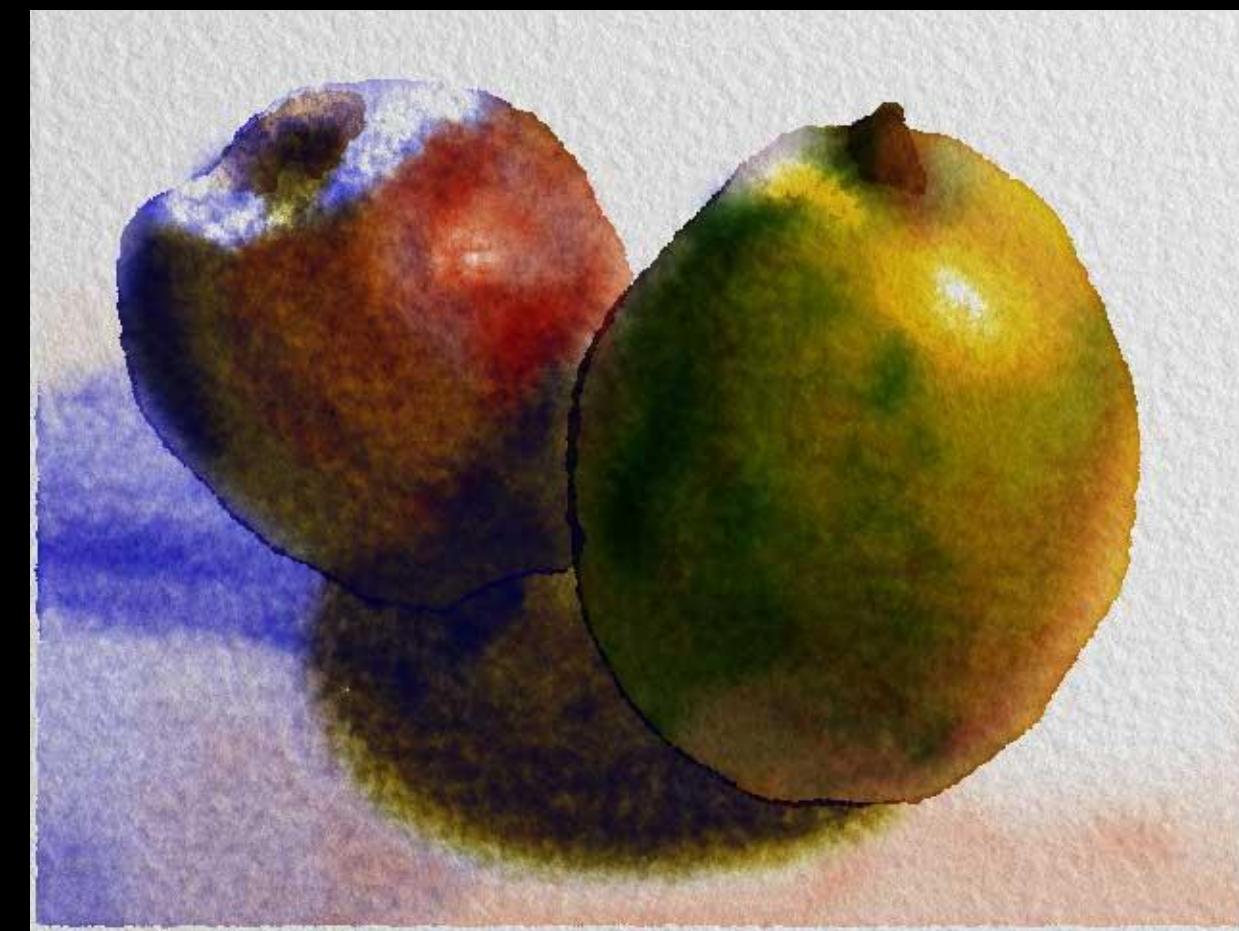
²Microsoft Research

³University of Washington

Image Analogies



A



A'



B



B'



Image Analogies

Goal: Process an image by example



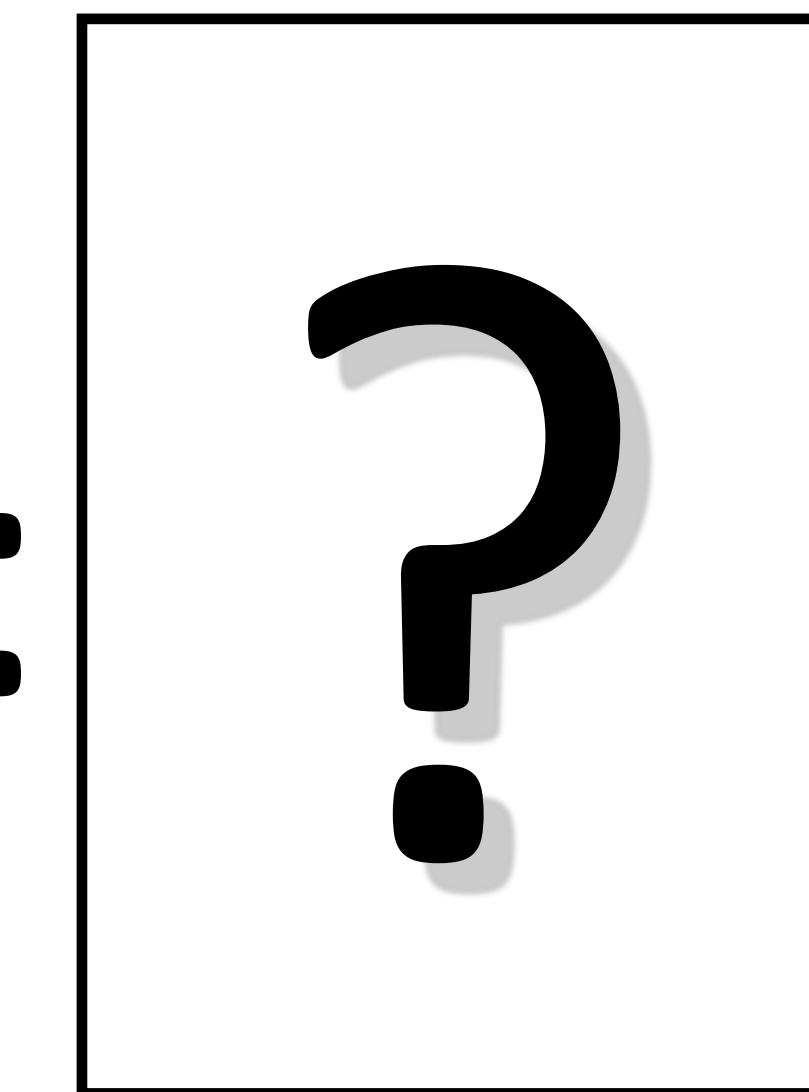
A



A'



B



B'

Non-parametric sampling



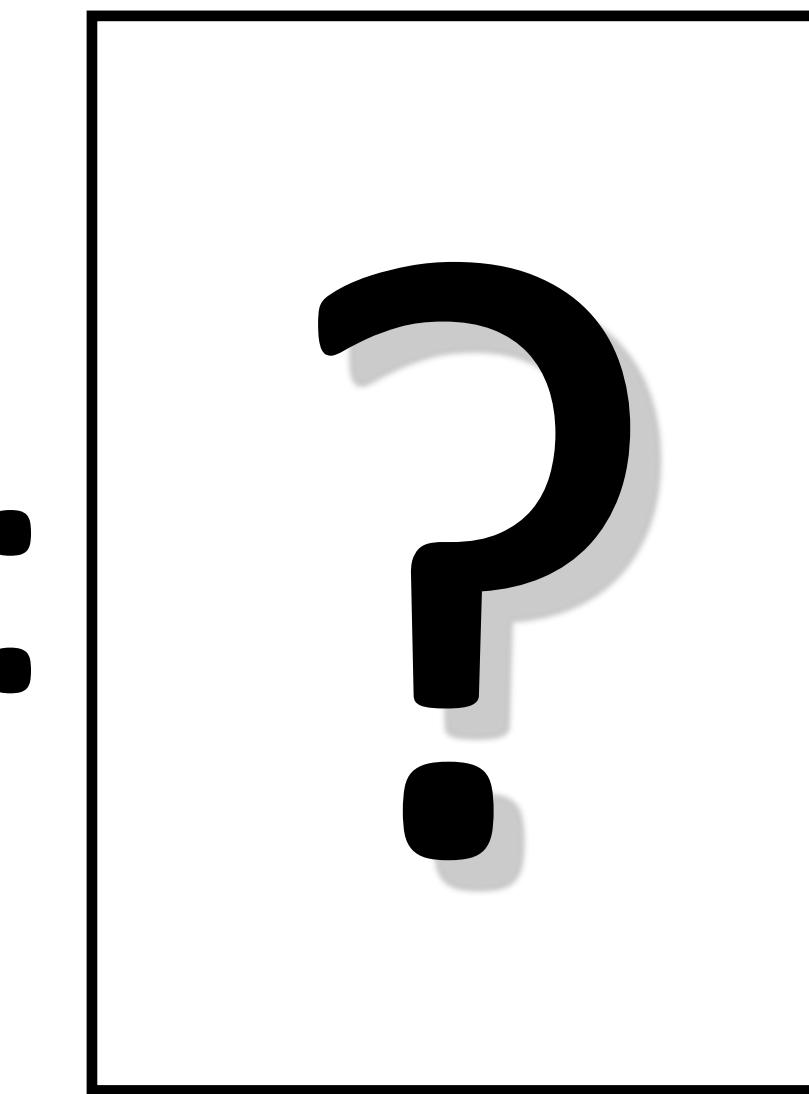
A



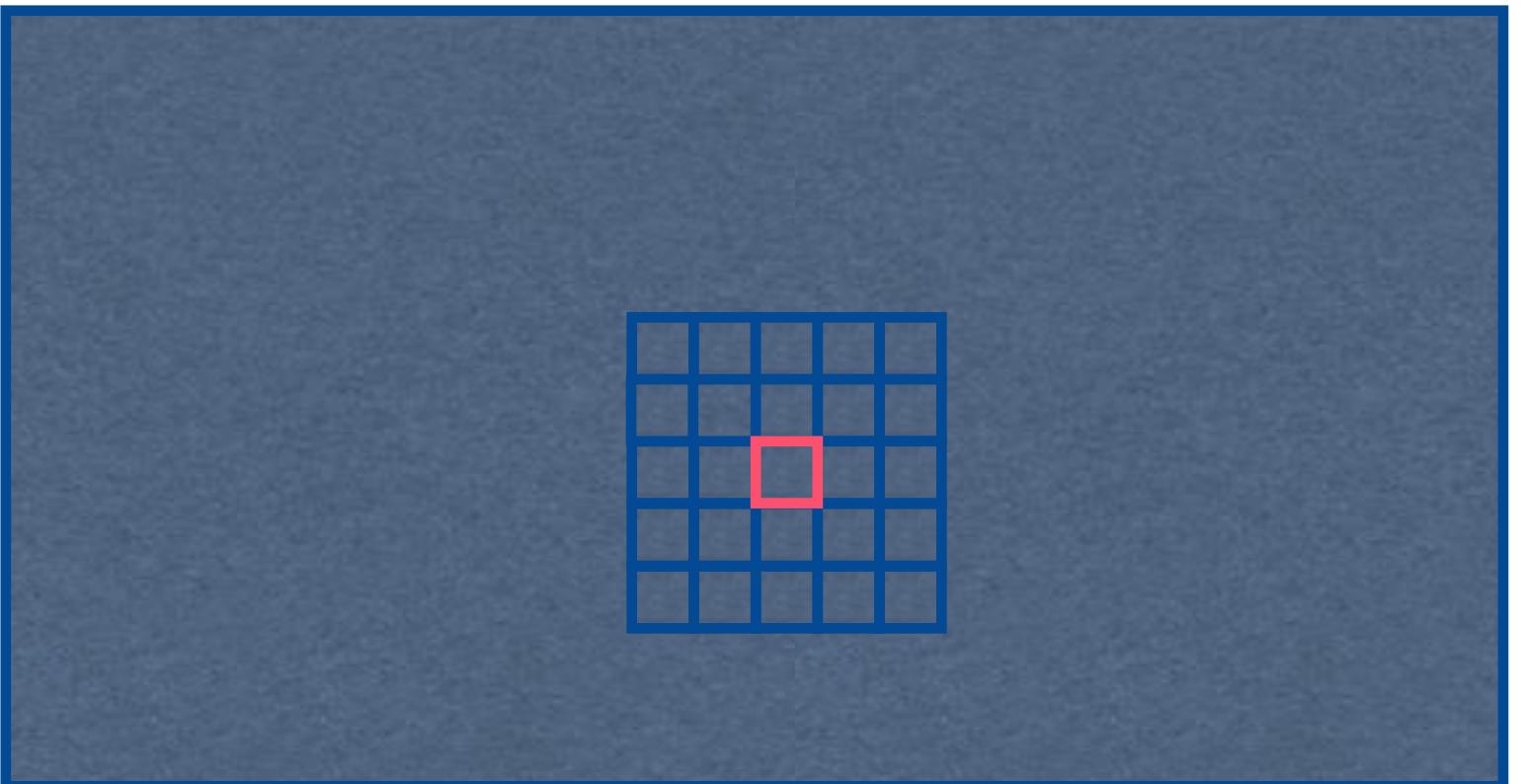
A'



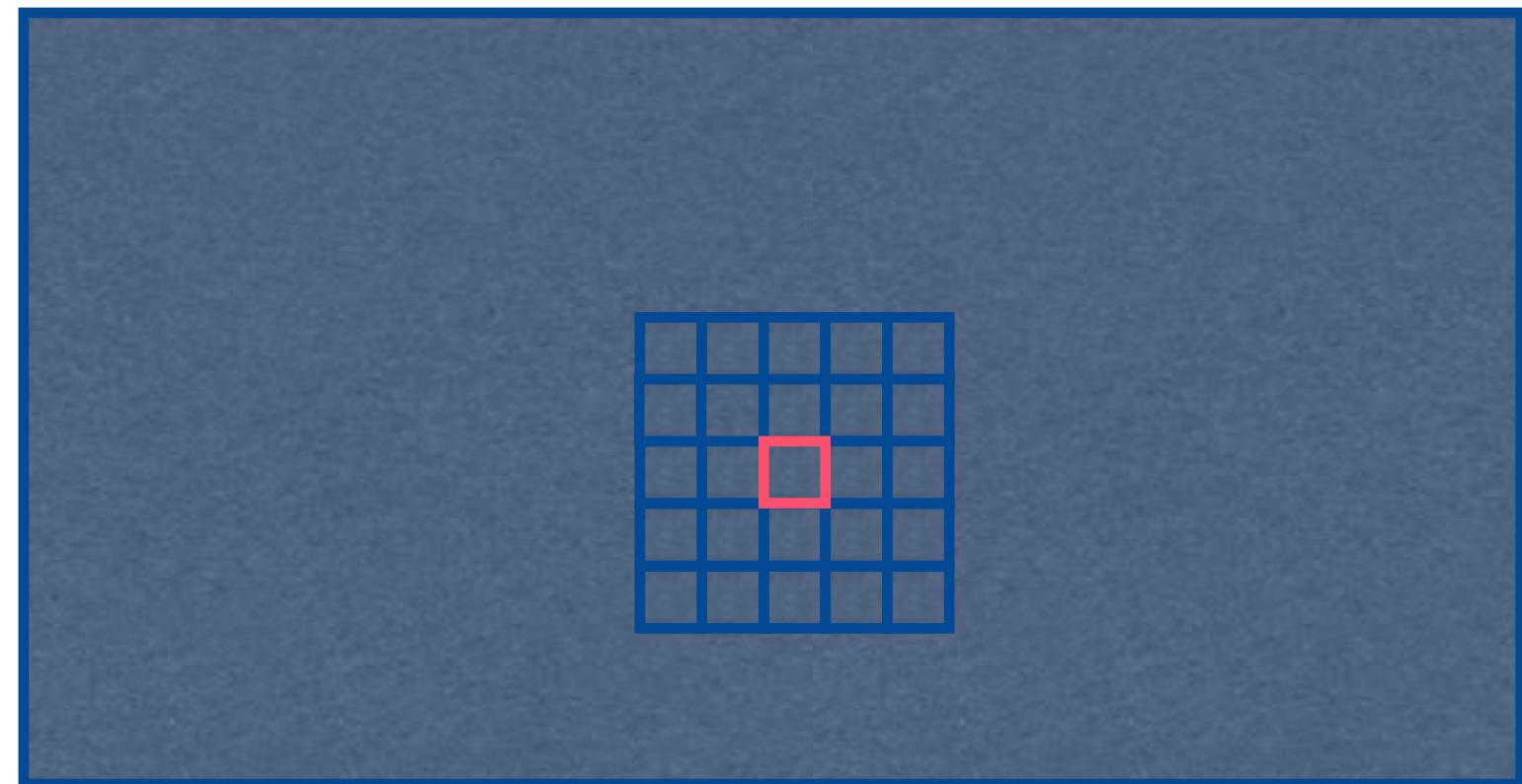
B



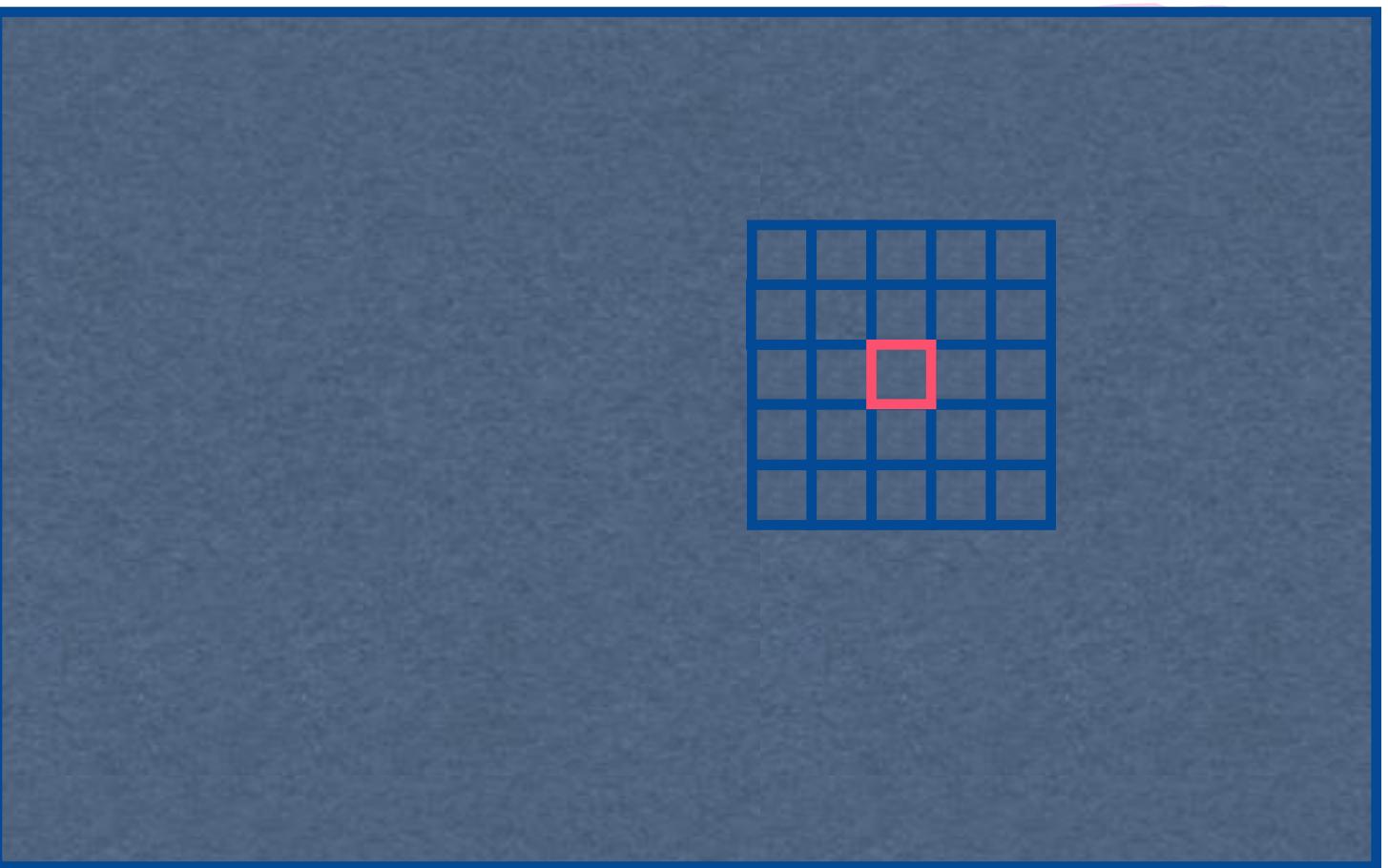
B'



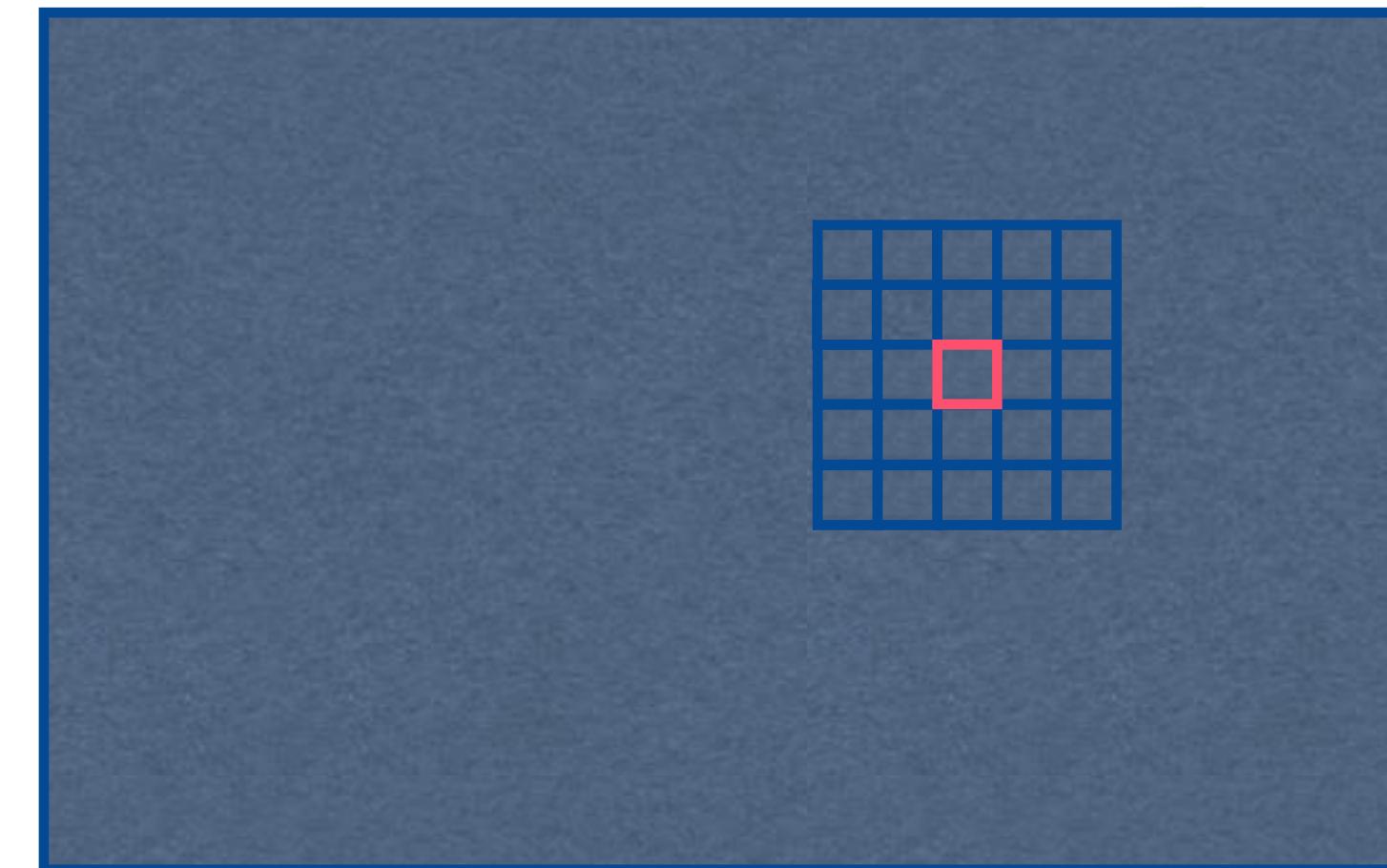
A



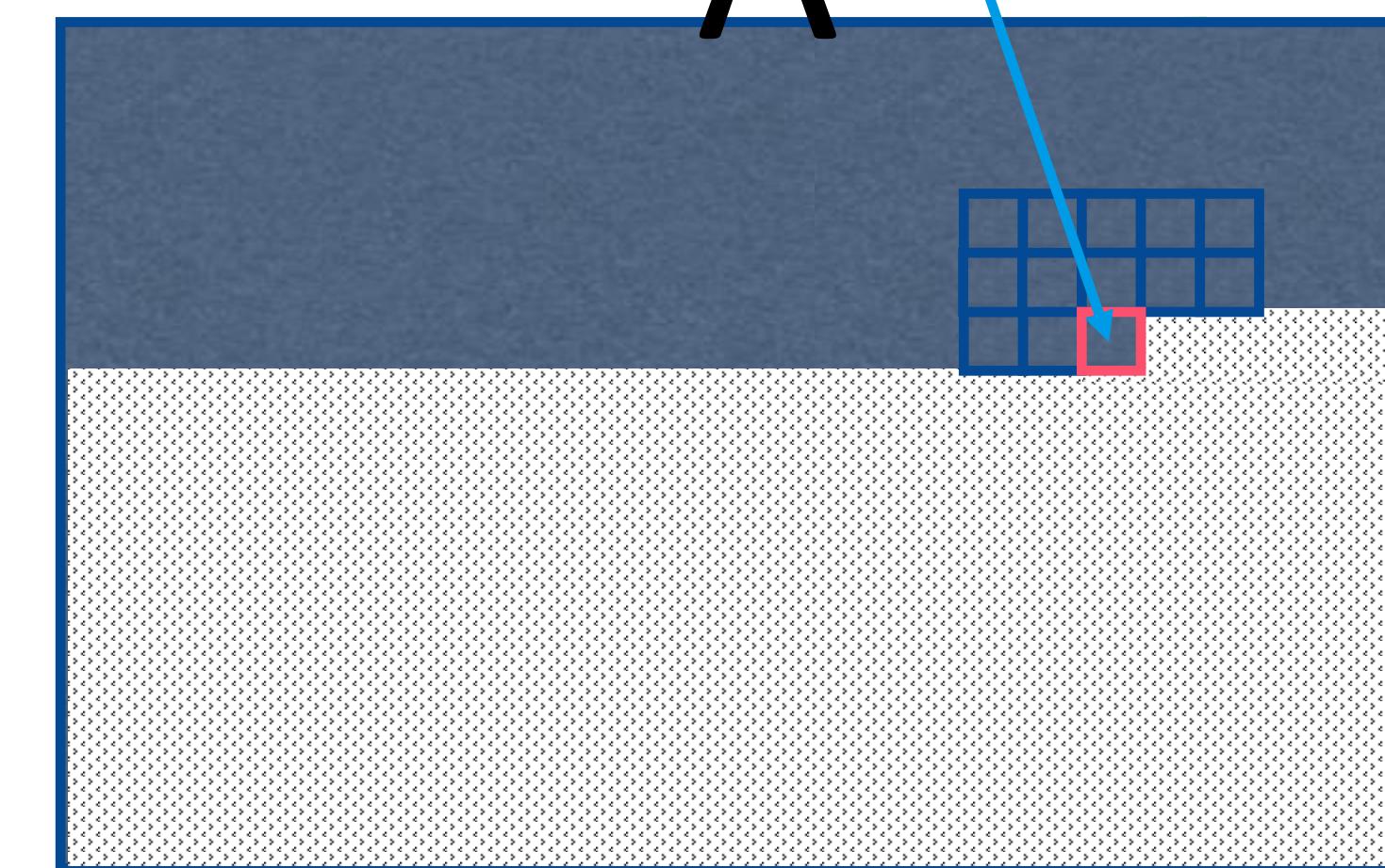
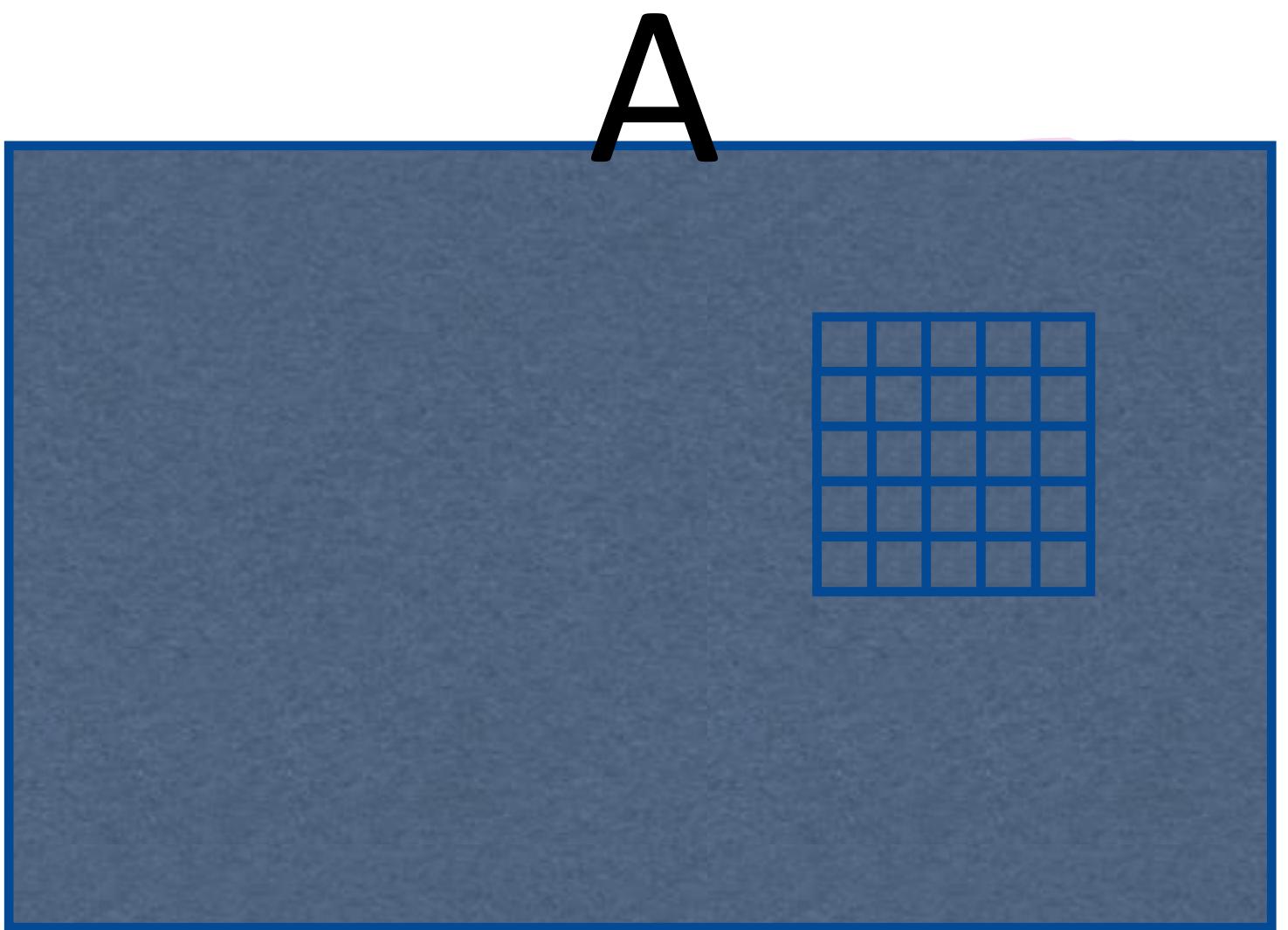
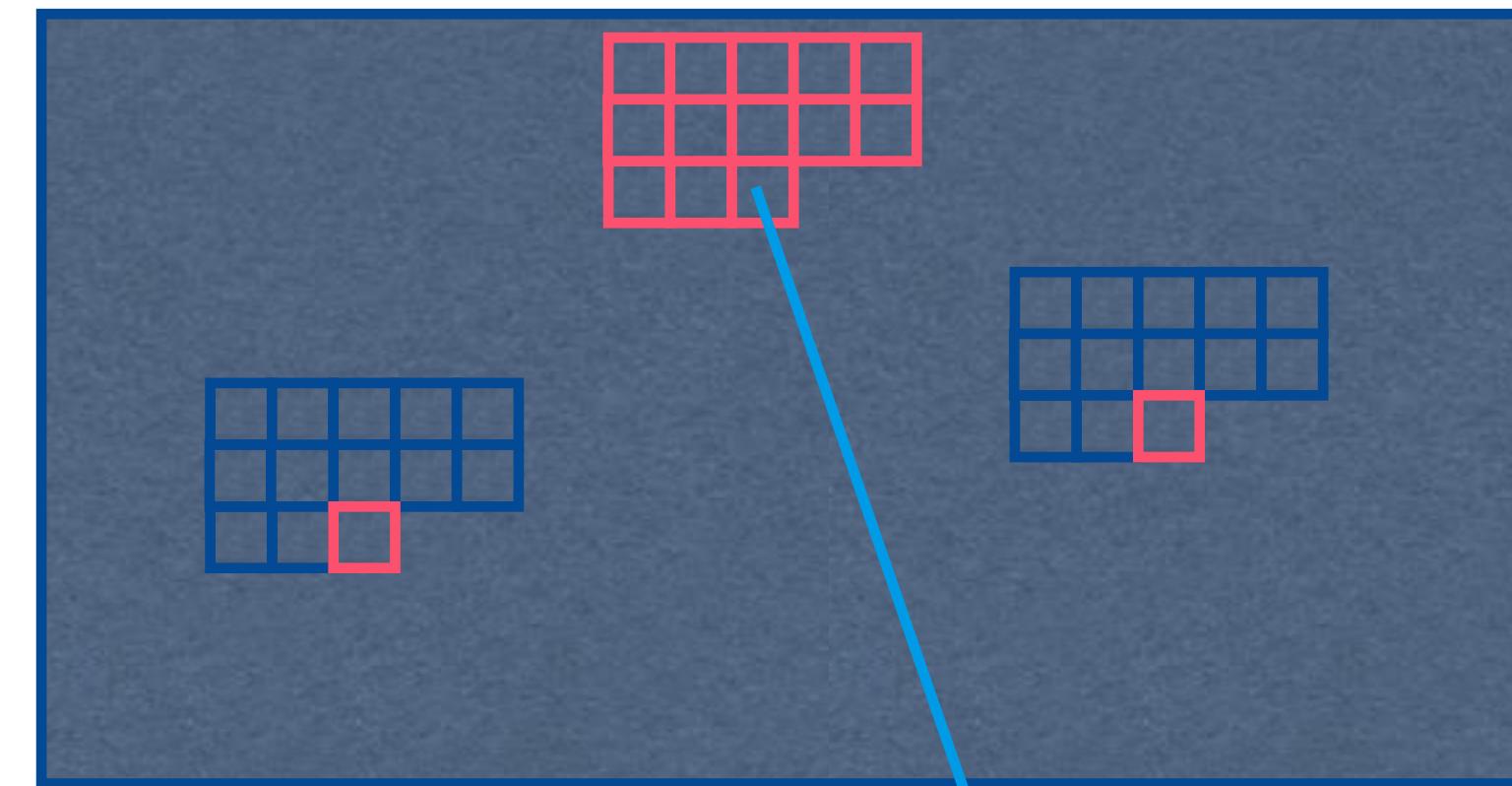
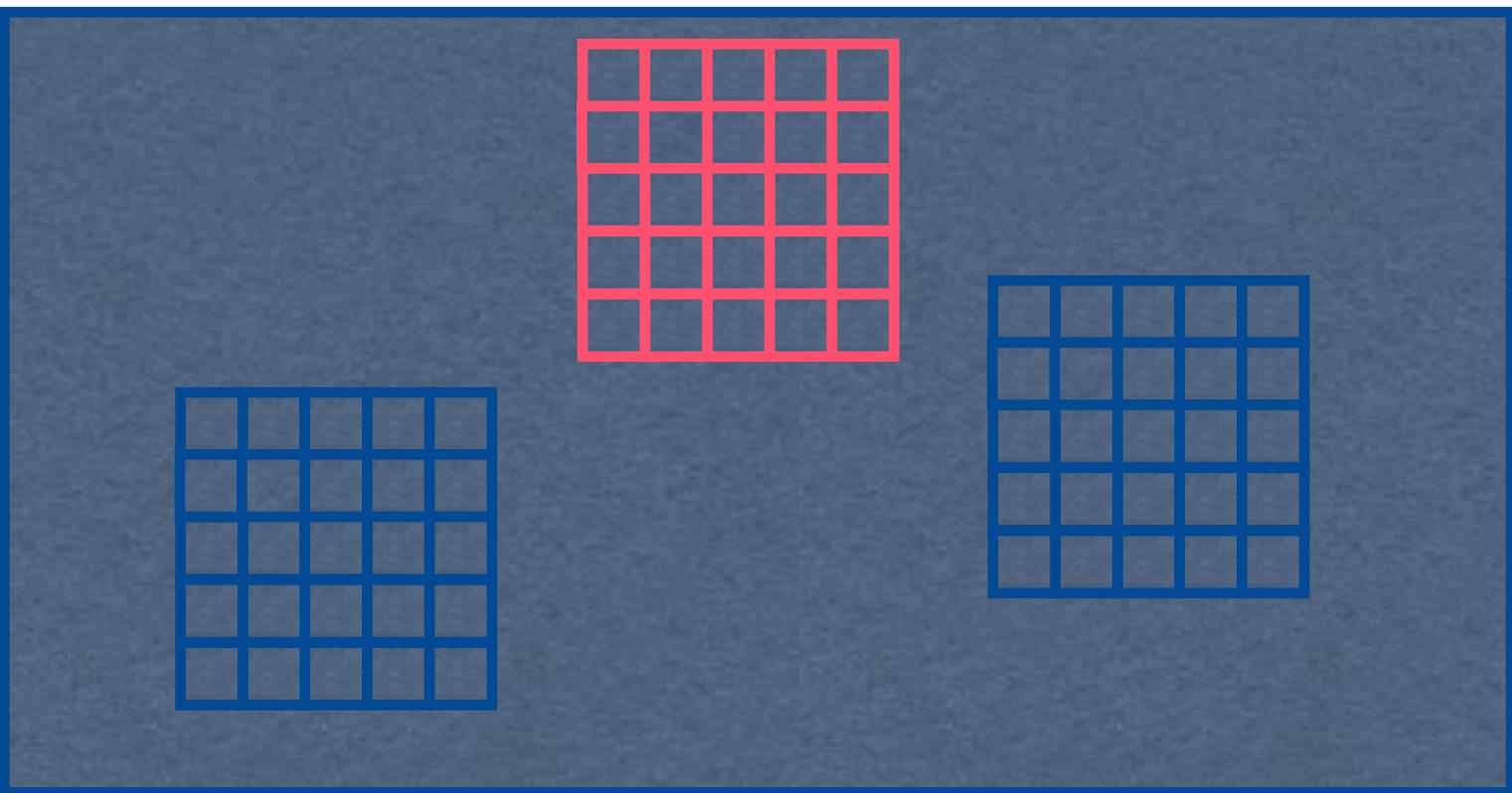
A'



B



B'



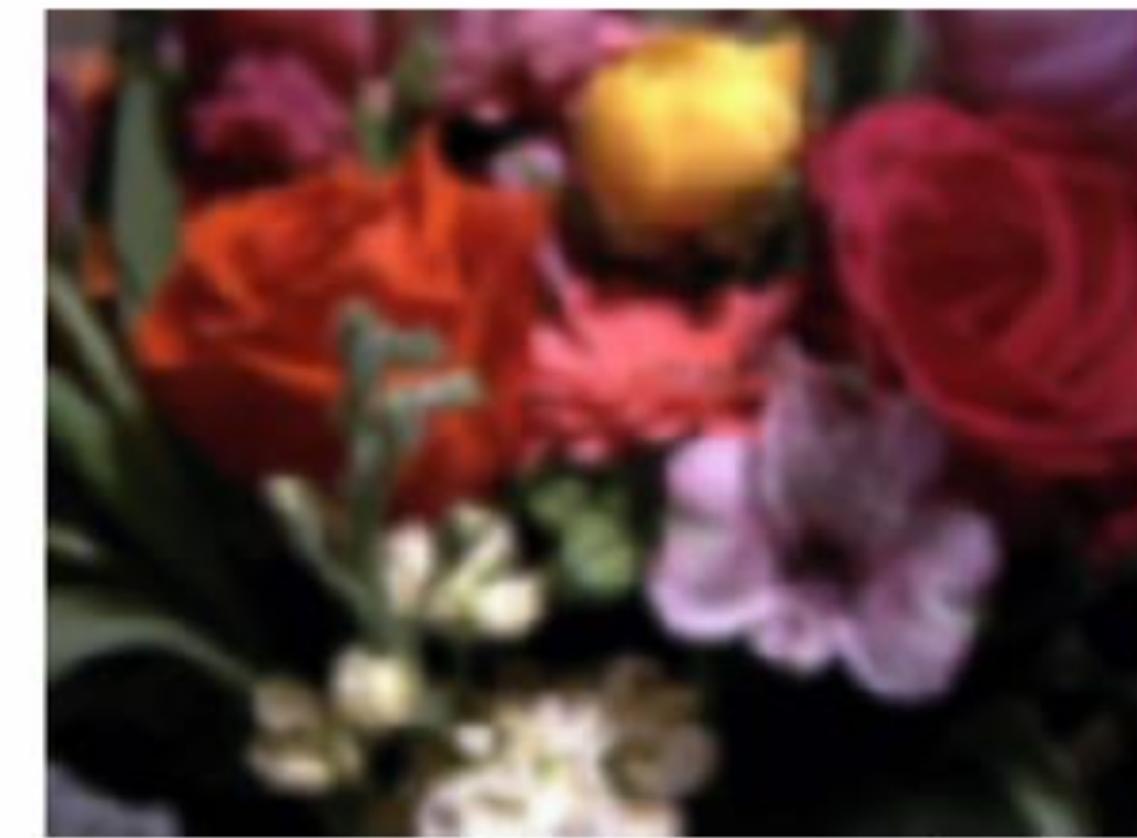
B

B'

Blur Filter



Unfiltered source (A)



Filtered source (A')



Unfiltered target (B)

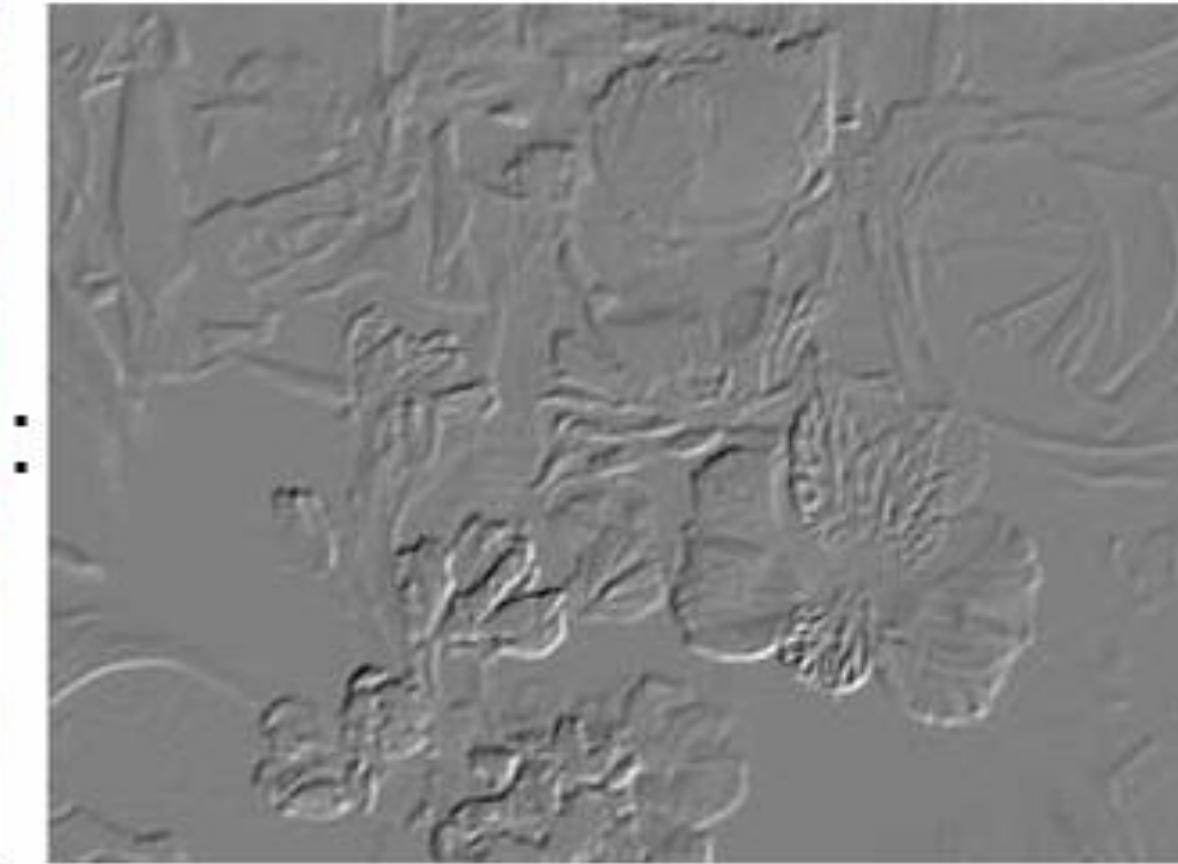


Filtered target (B')

Edge Filter



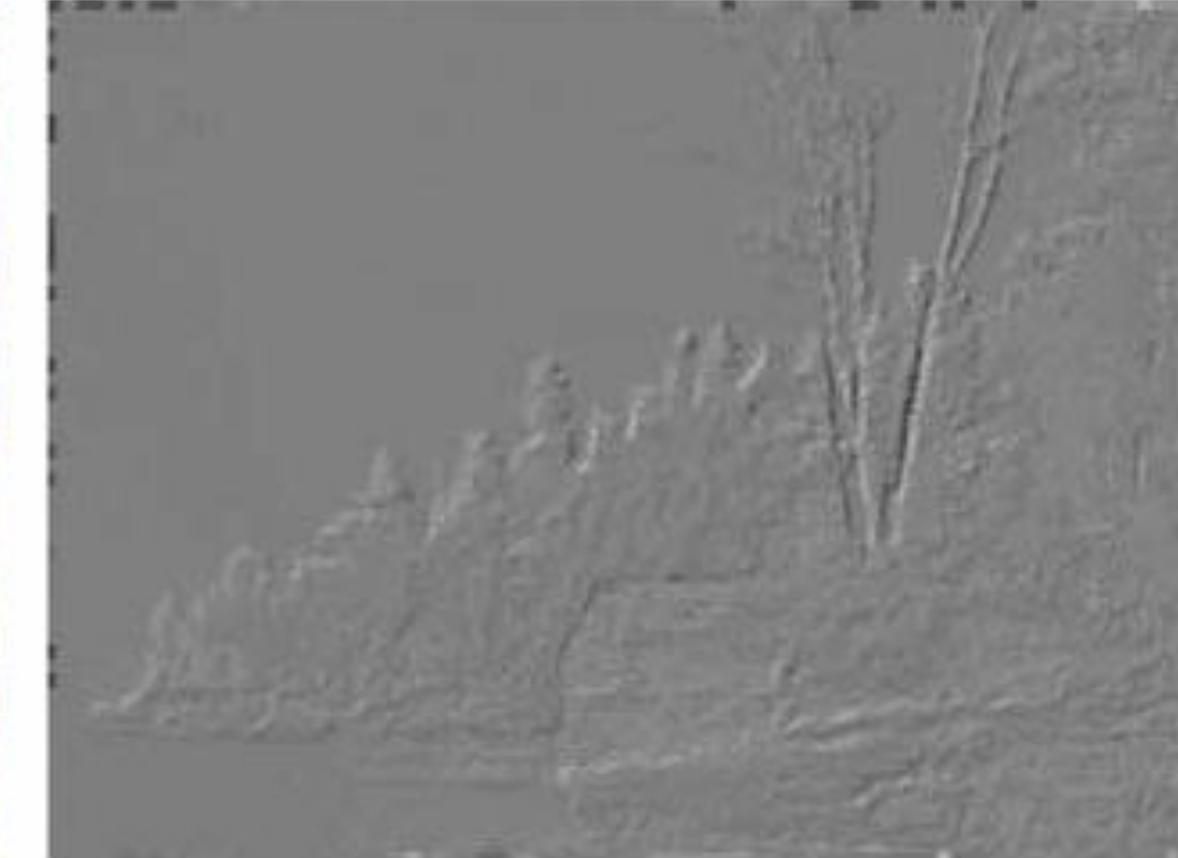
Unfiltered source (*A*)



Filtered source (*A'*)



Unfiltered target (*B*)



Filtered target (*B'*)

Artistic Filters



A



A'



B



B'

Colorization



Unfiltered source (*A*)



Filtered source (*A'*)



Unfiltered target (*B*)

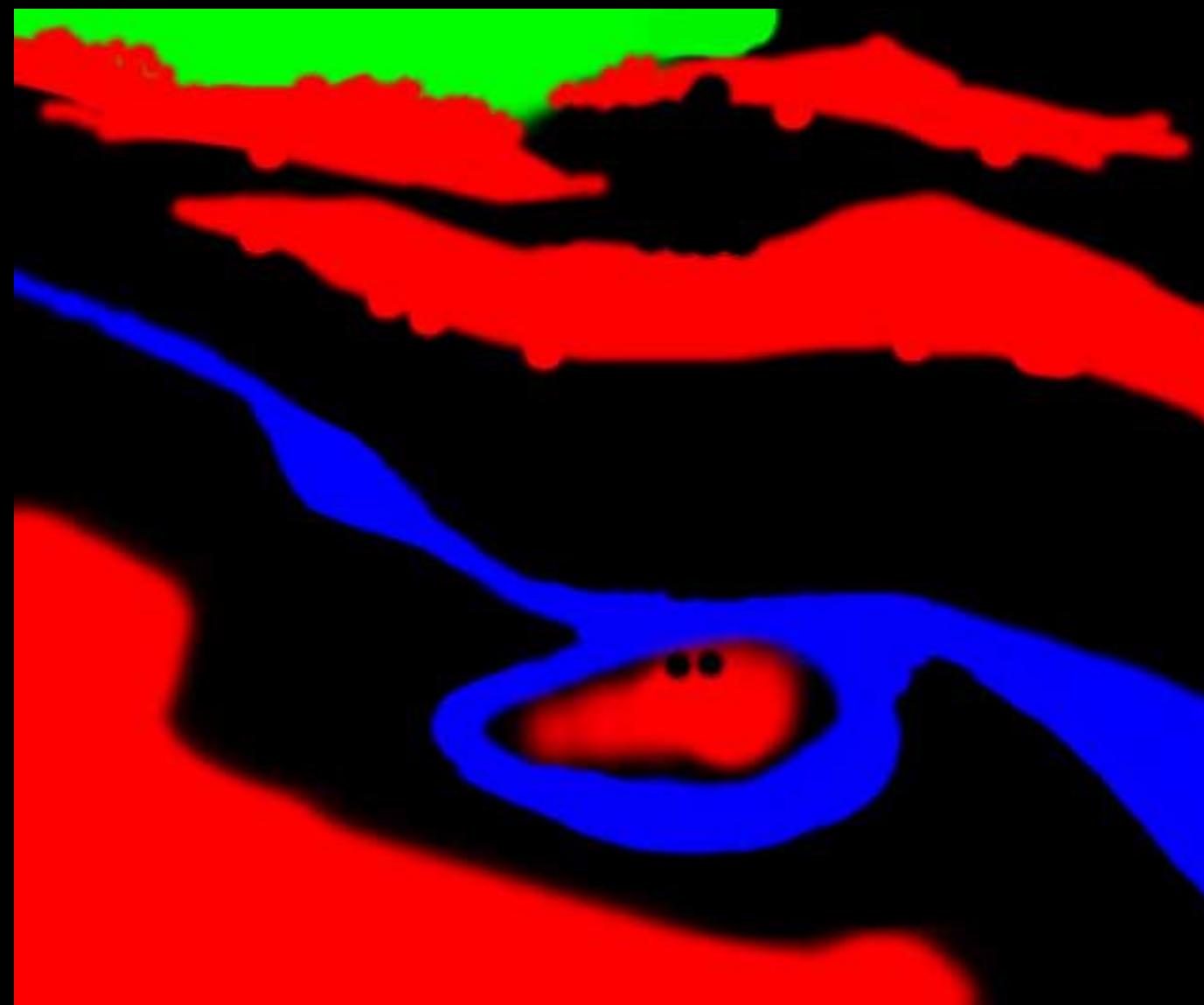


Filtered target (*B'*)

Texture-by-numbers



A



B



A'



B'

Visual Prompting via Image Inpainting

Amir Bar*, Yossi Gandelsman*,

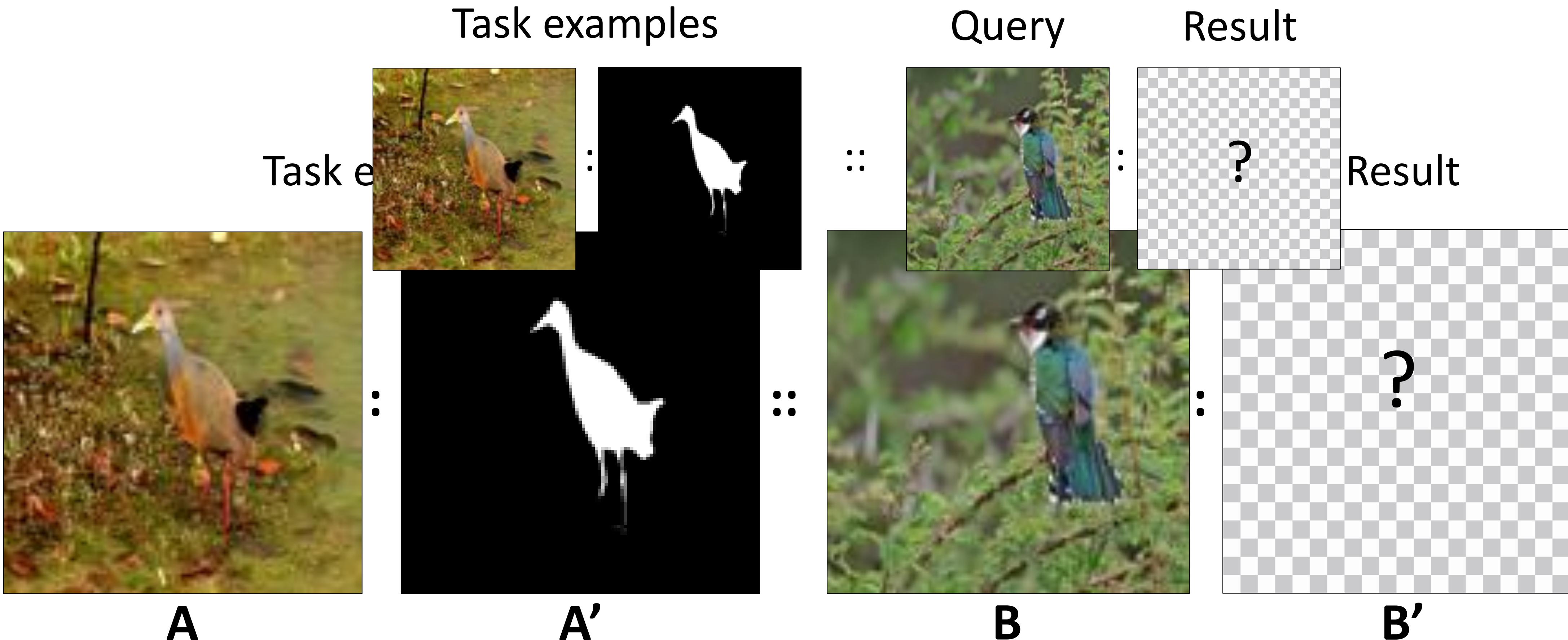
Trevor Darrell, Amir Globerson, Alexei A Efros

NeurIPS 2022



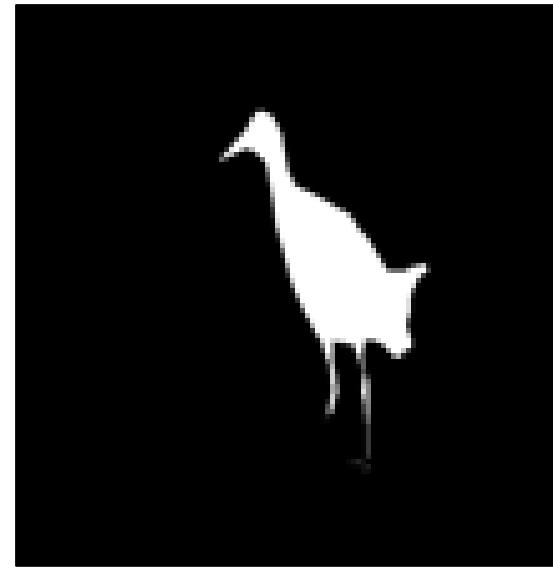
* Equal contribution

Visual Prompting

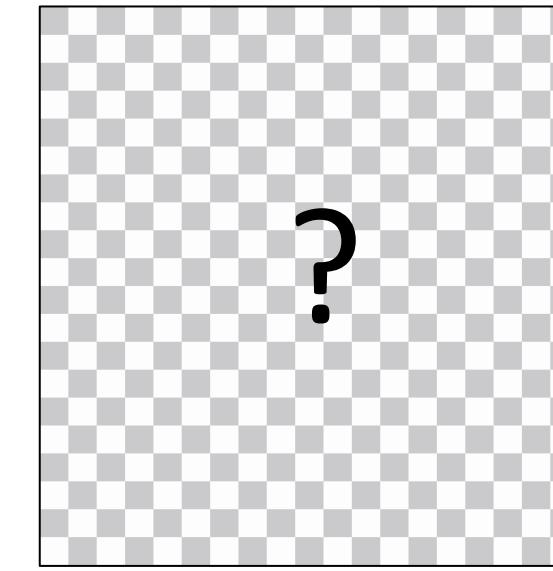
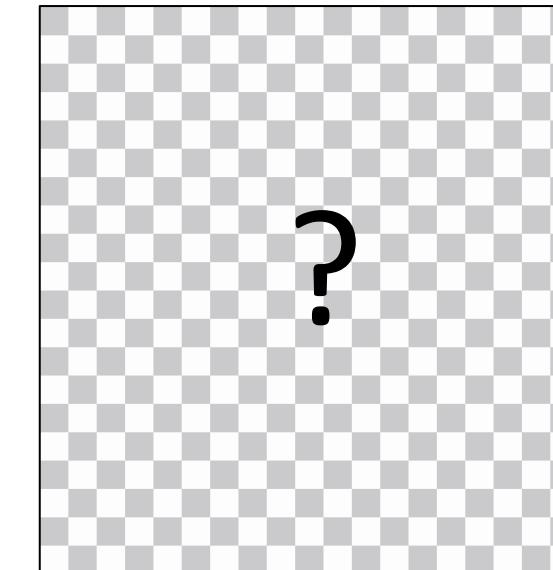
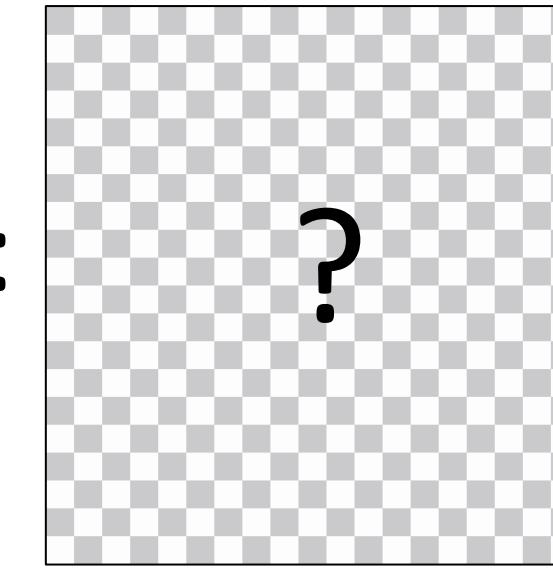


Visual Prompting

Task examples



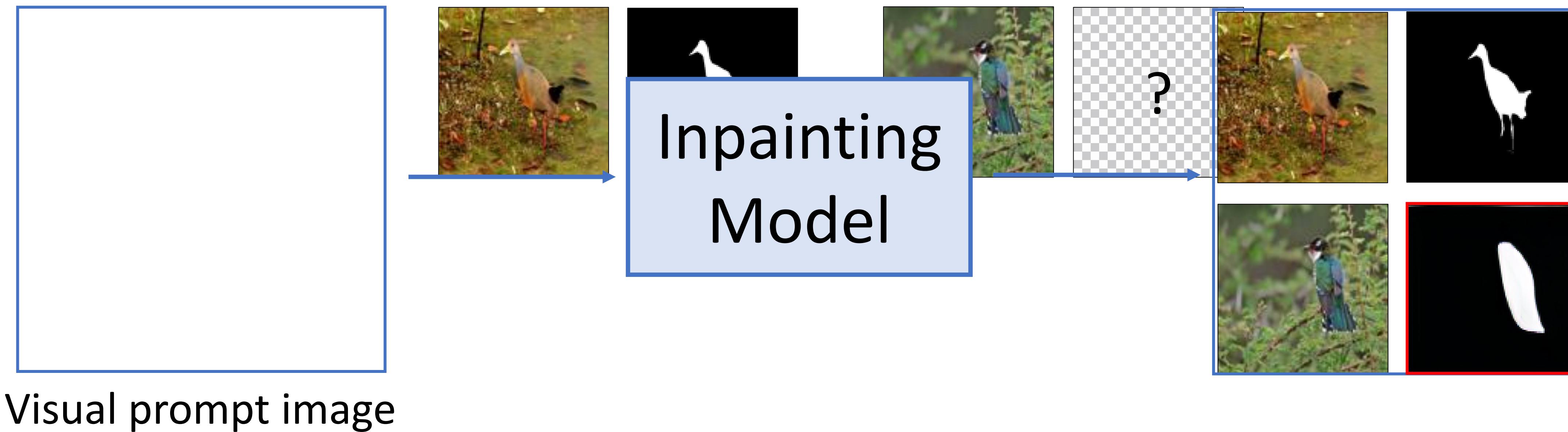
::



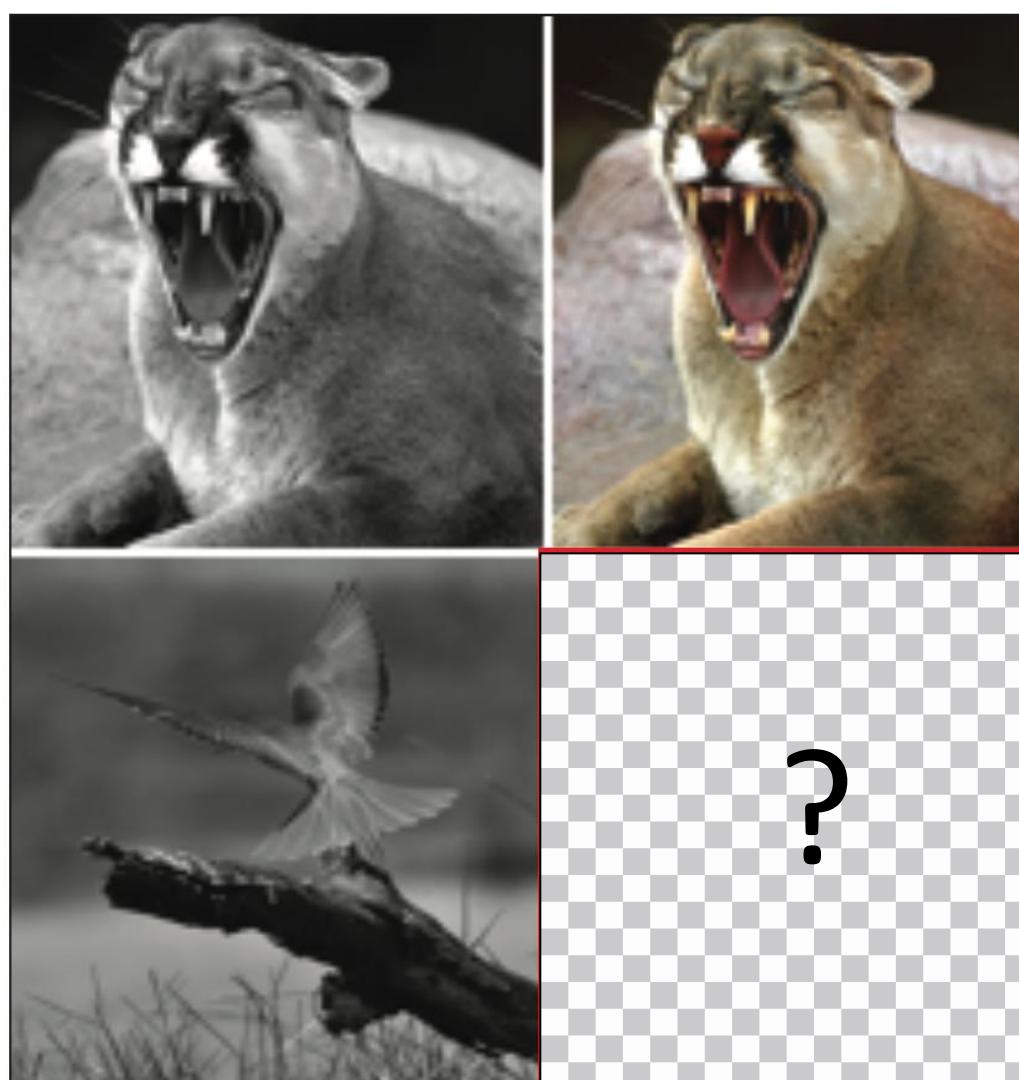
Query

Result

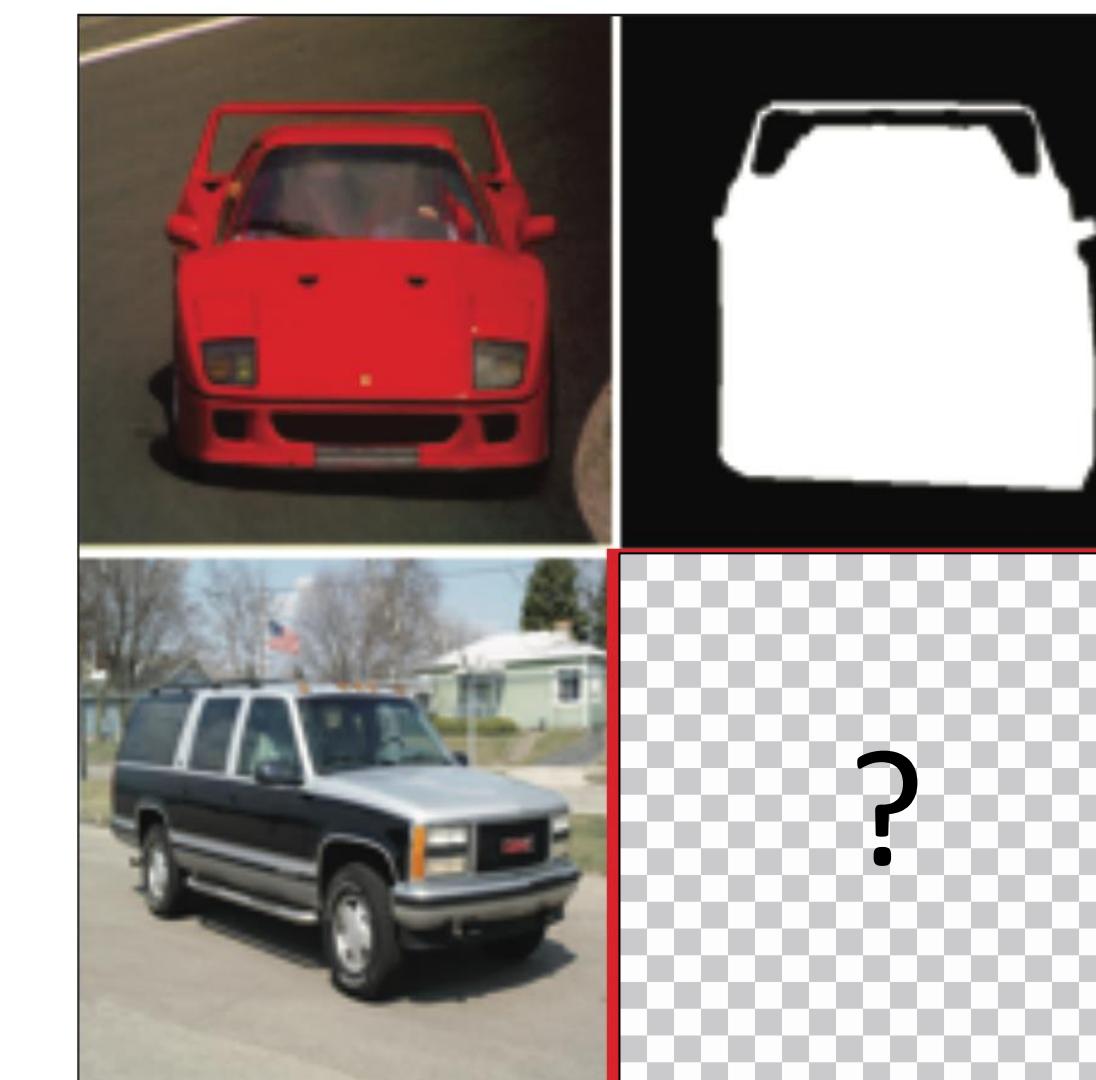
Inpainting models to the rescue!



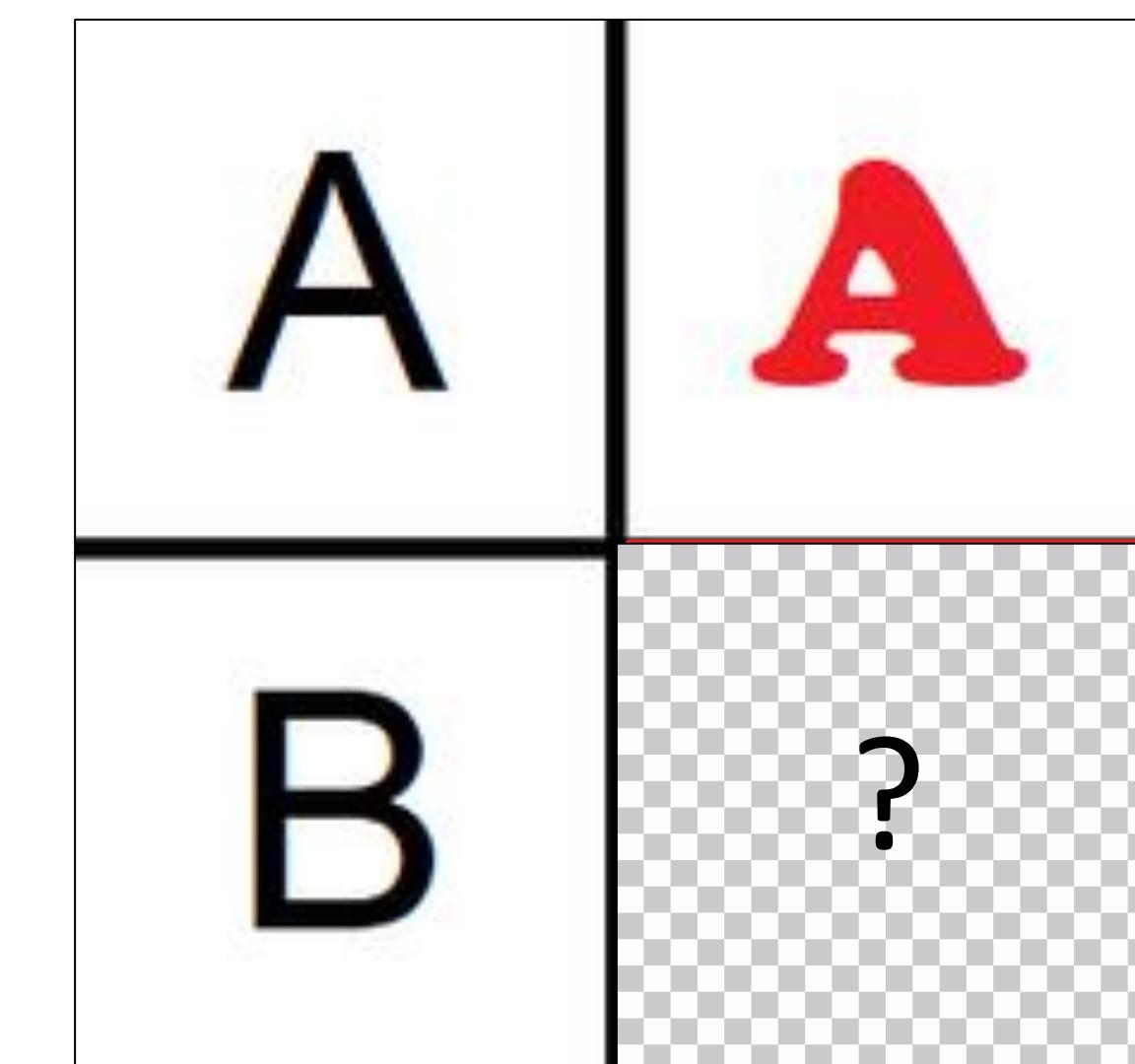
Wide range of tasks



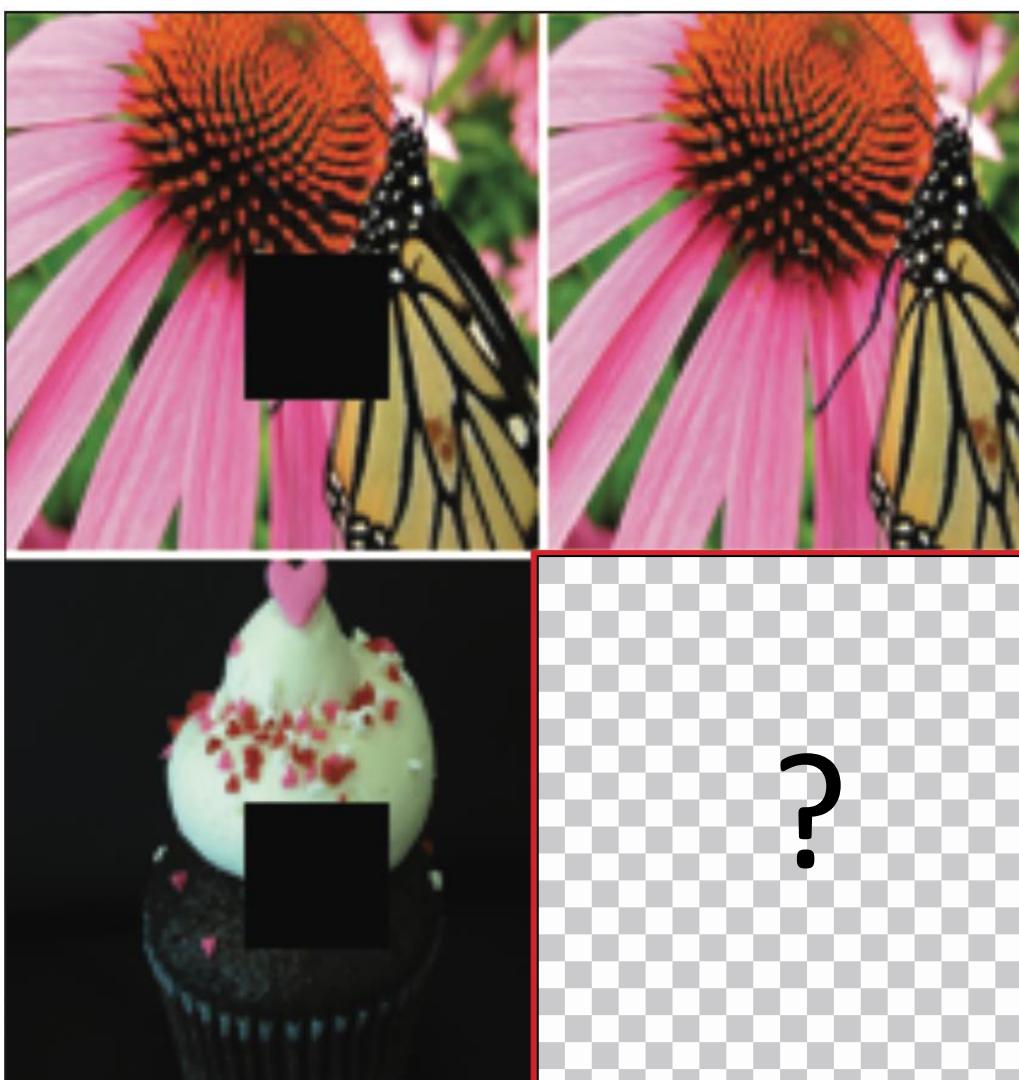
Colorization



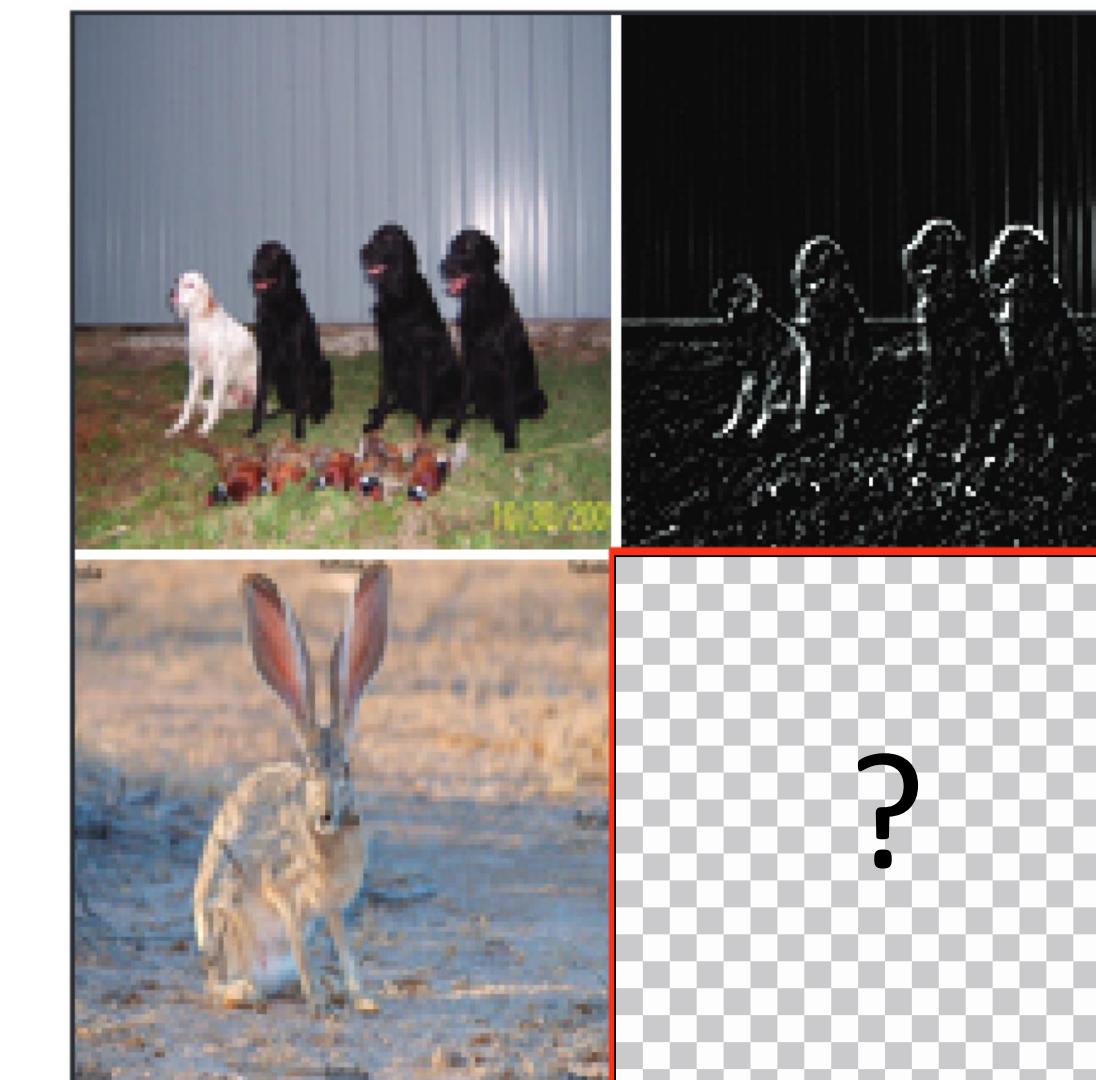
Segmentation



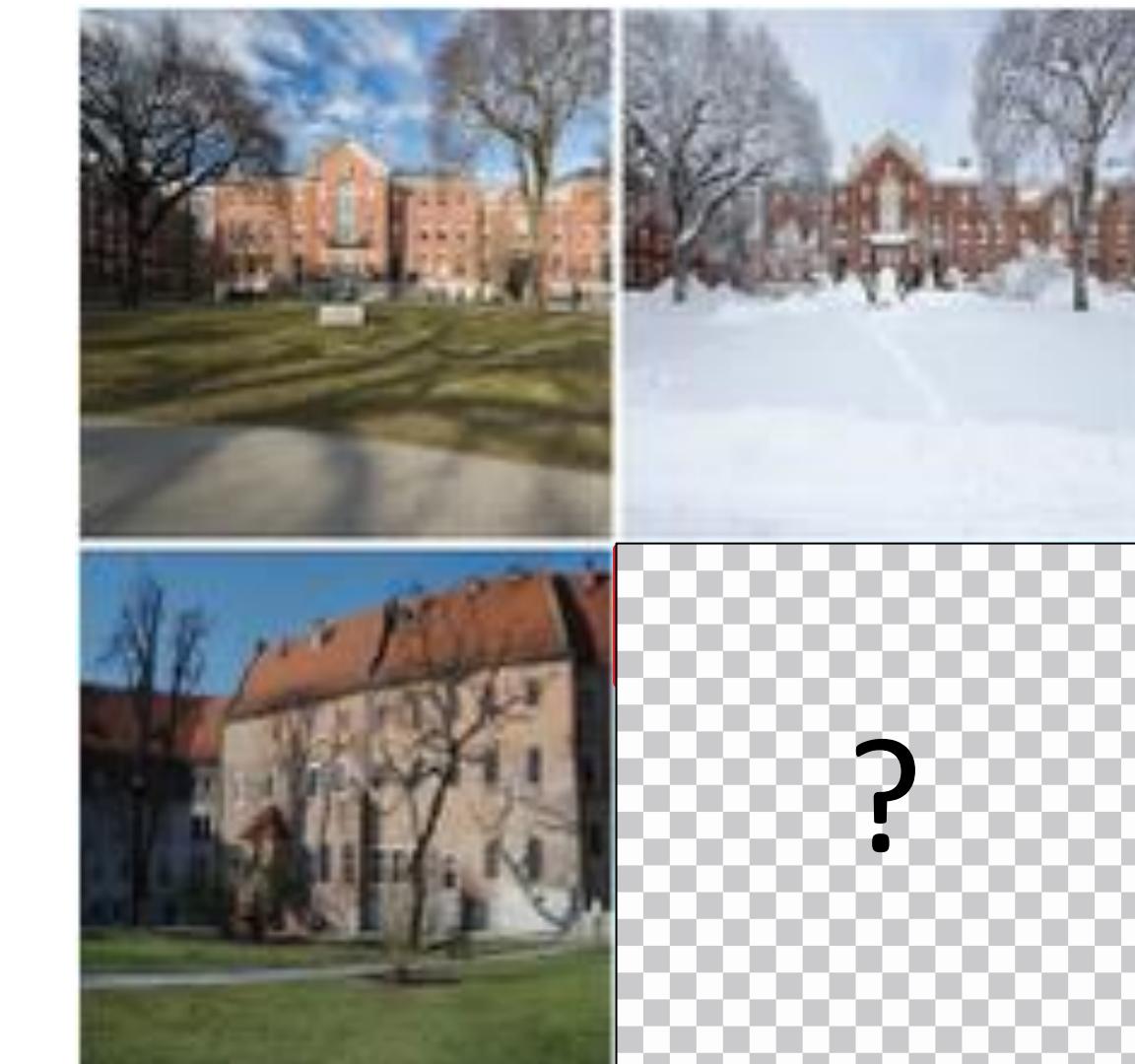
Font Style Transfer



Inpainting

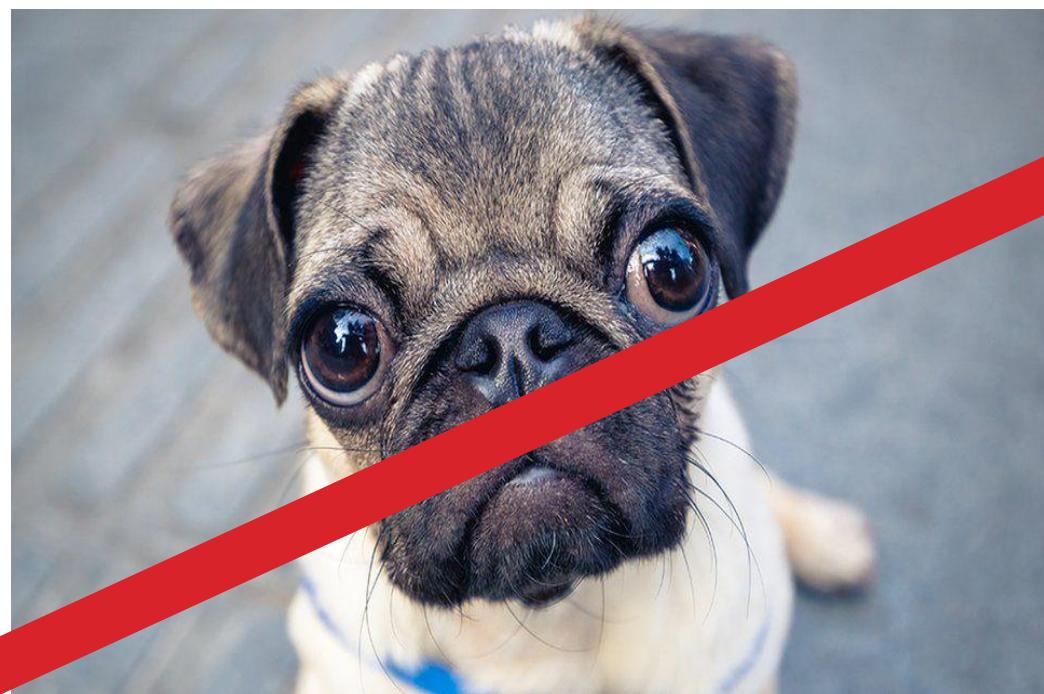


Edge Detection

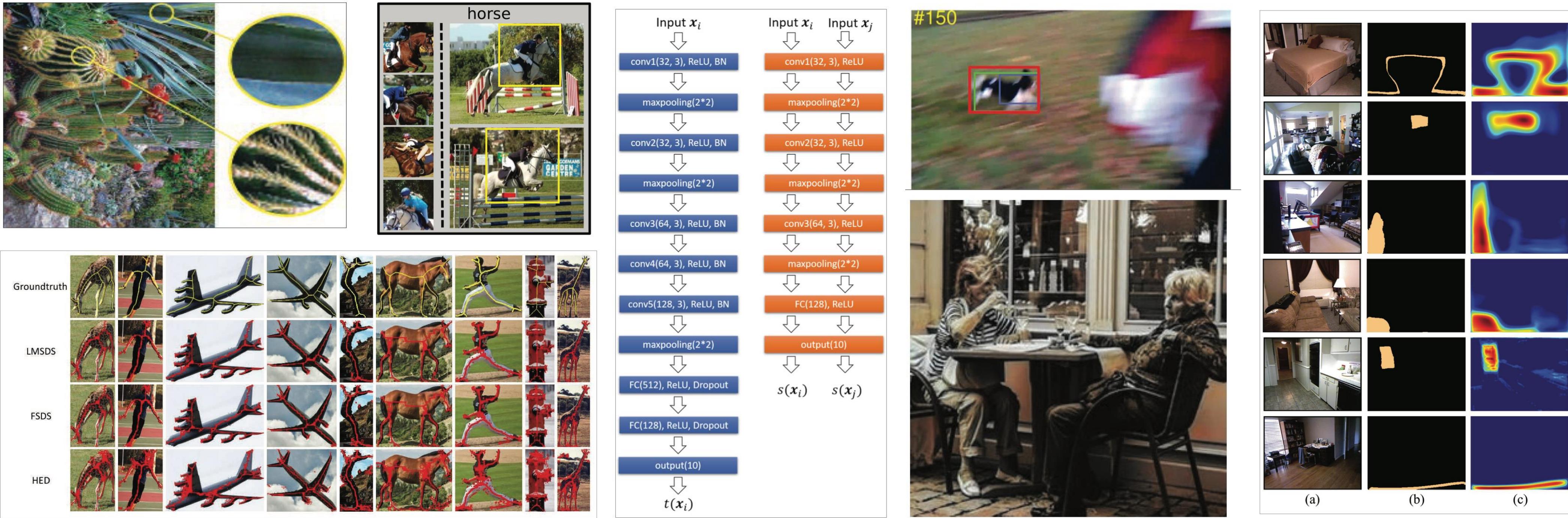


Style Transfer

Training Data

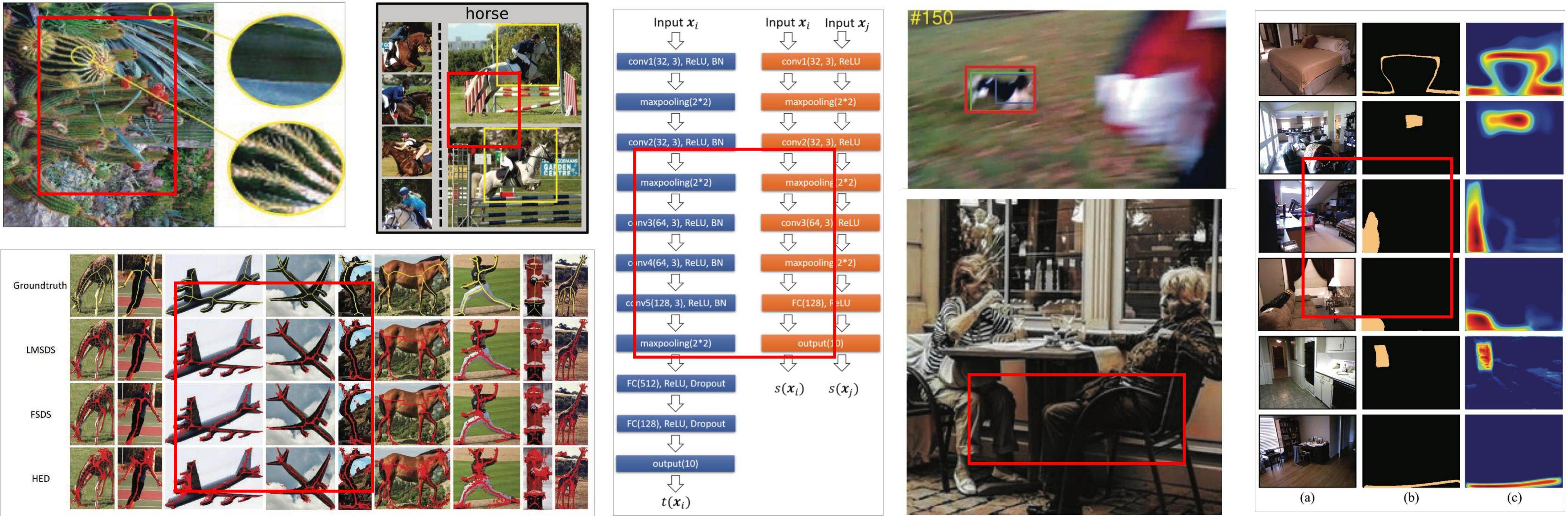


Computer Vision Figures Dataset (CVFD)



- 88k images from cs.CV arxiv papers (2010 to 2022)
- Many figures have grid-like structure

Training Time

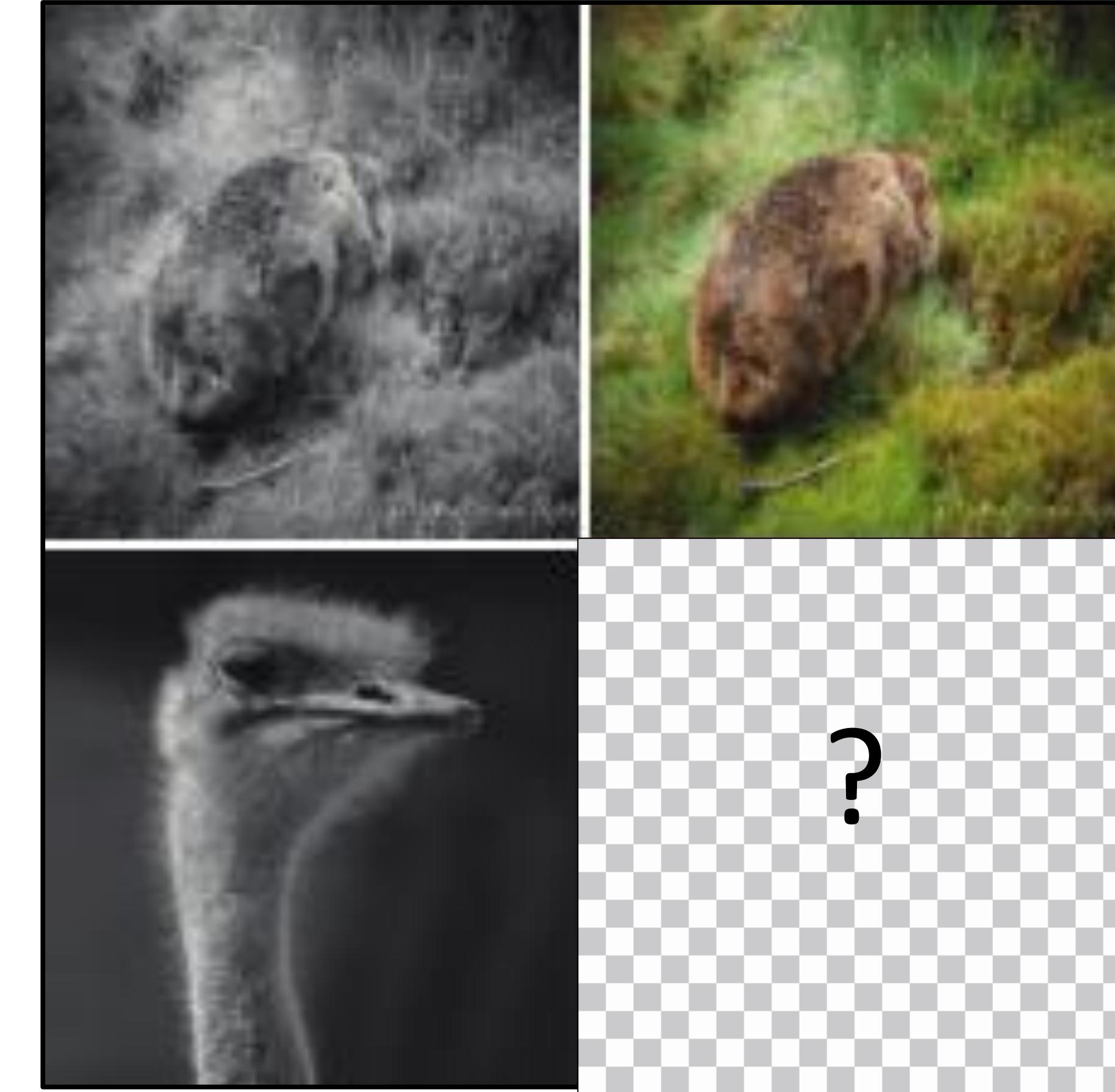


- Random 224x224 crops from figure images
 - No parsing
- We train MAE-VQGAN, a variant of Masked-Autoencoder (MAE)

Results

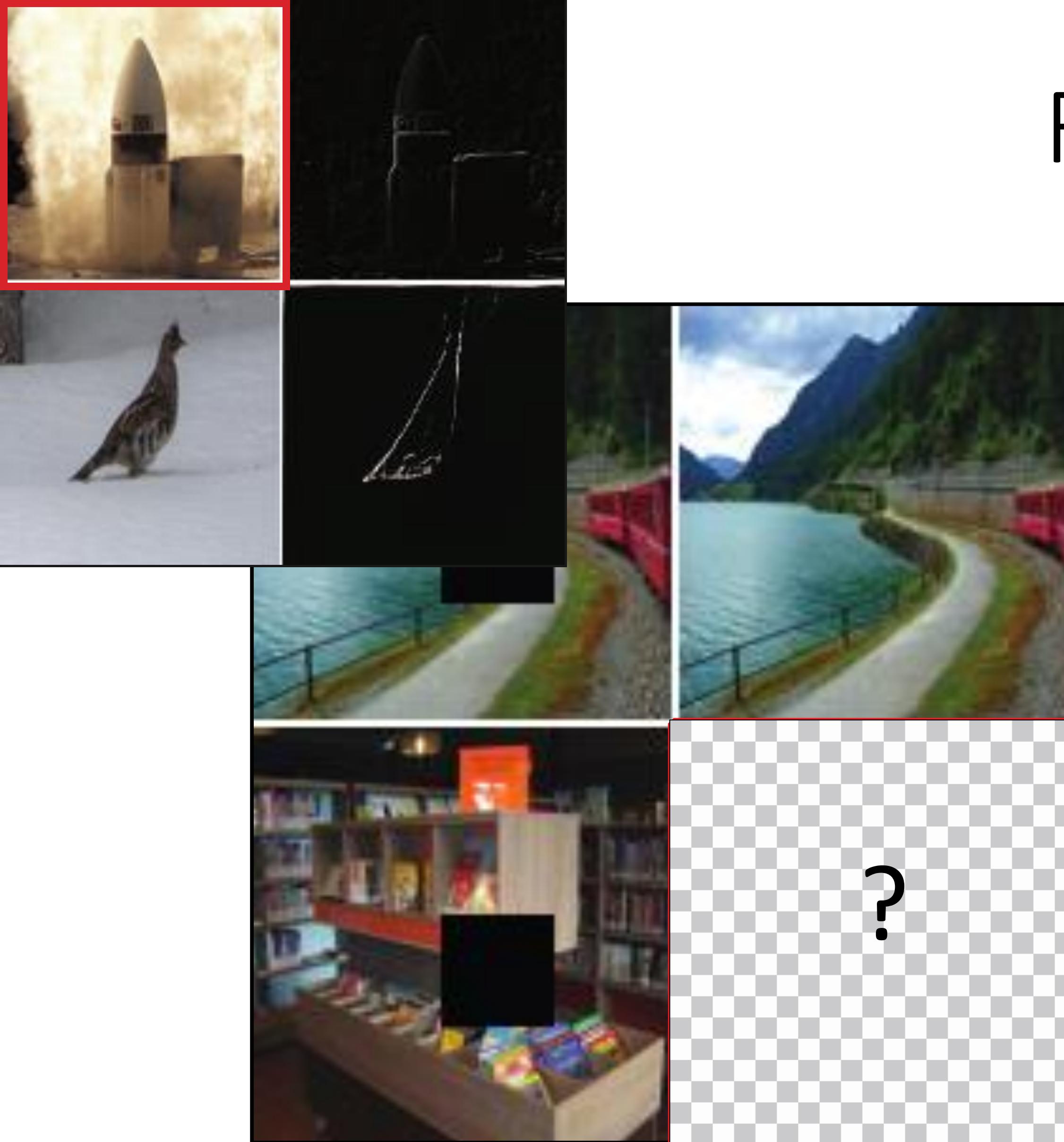


Segmentation

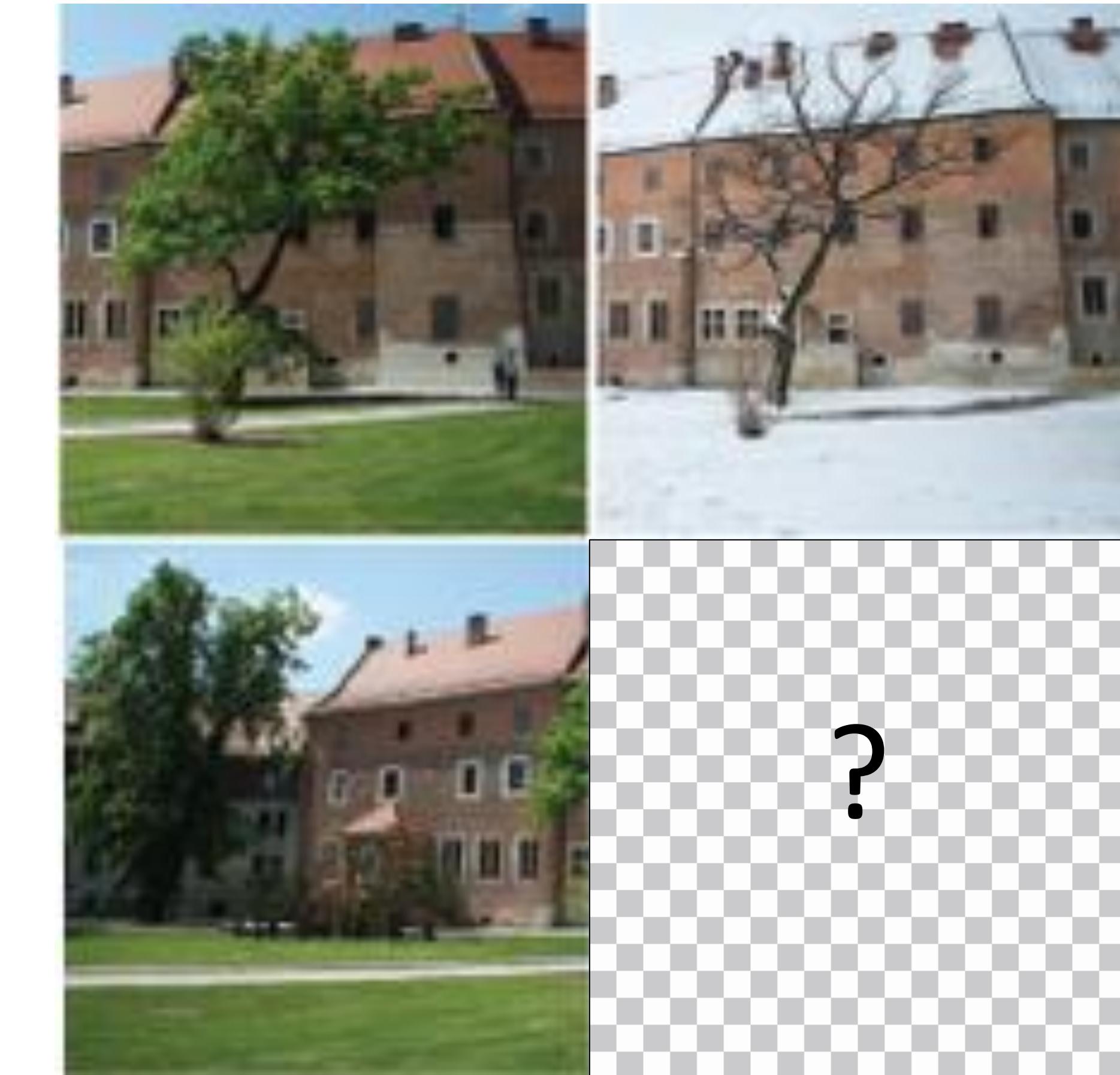


Colorization

Results

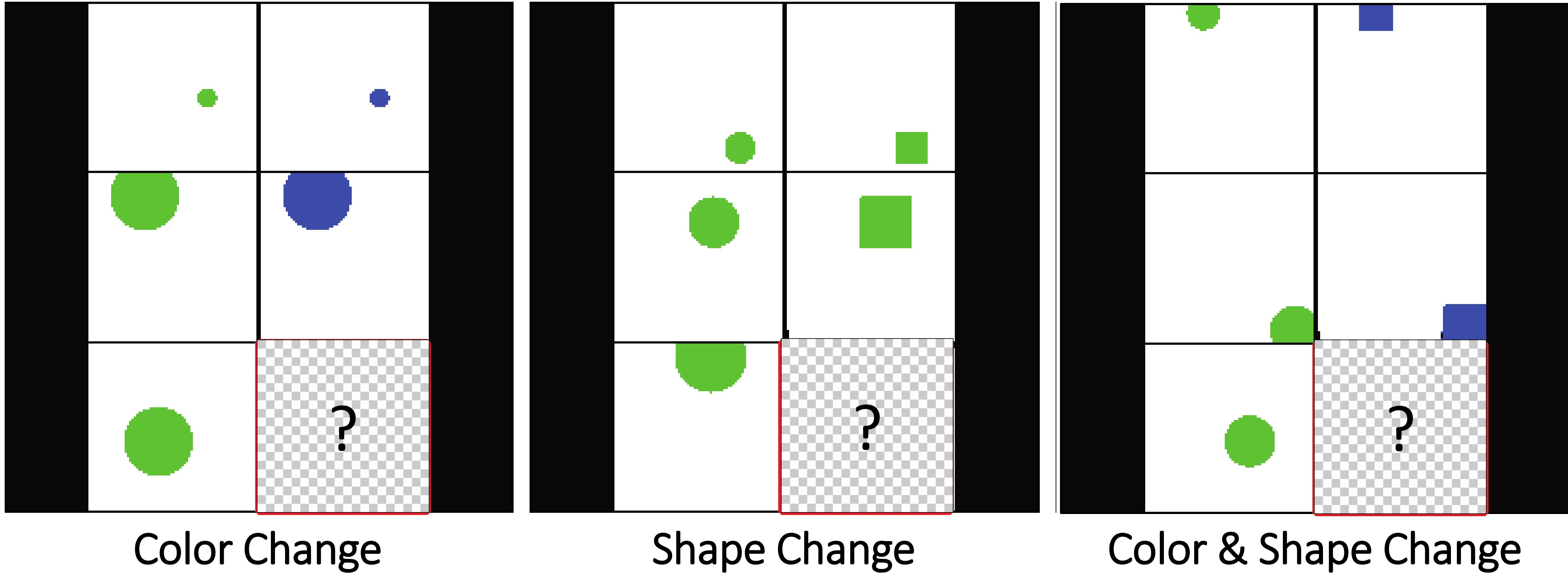


Inpainting



Style Transfer

Synthetic Experiments

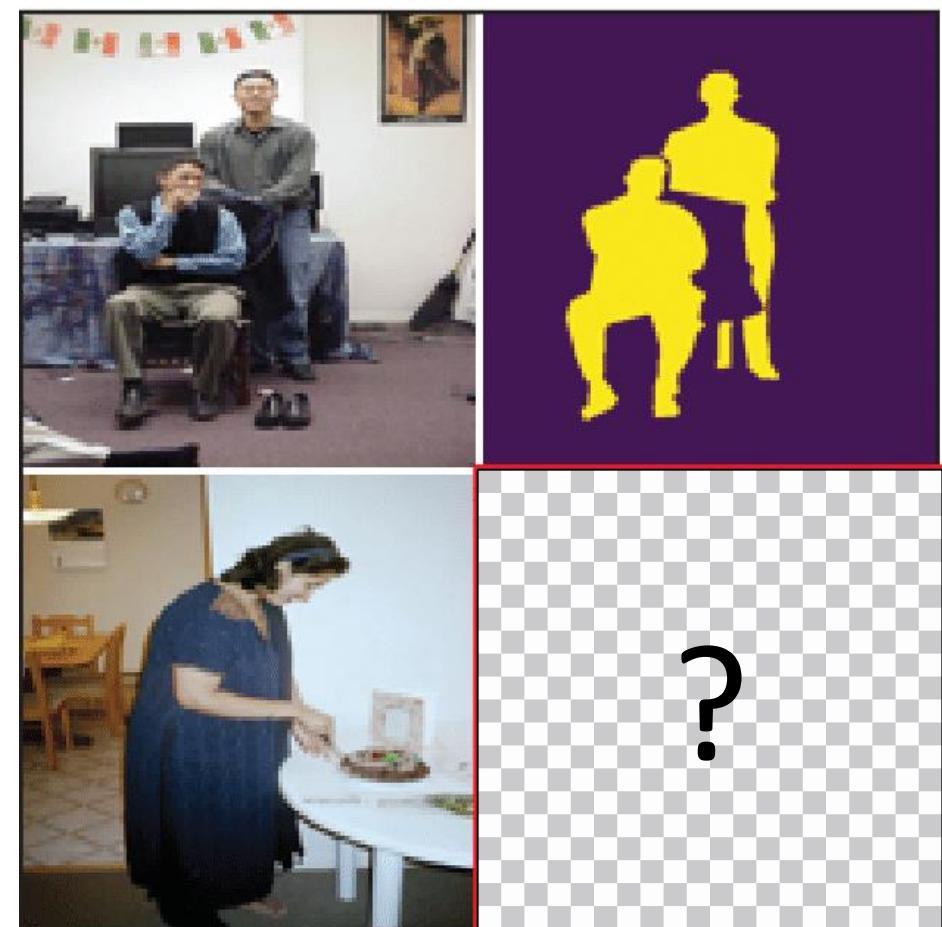


Prompt Variations: Input-output layout



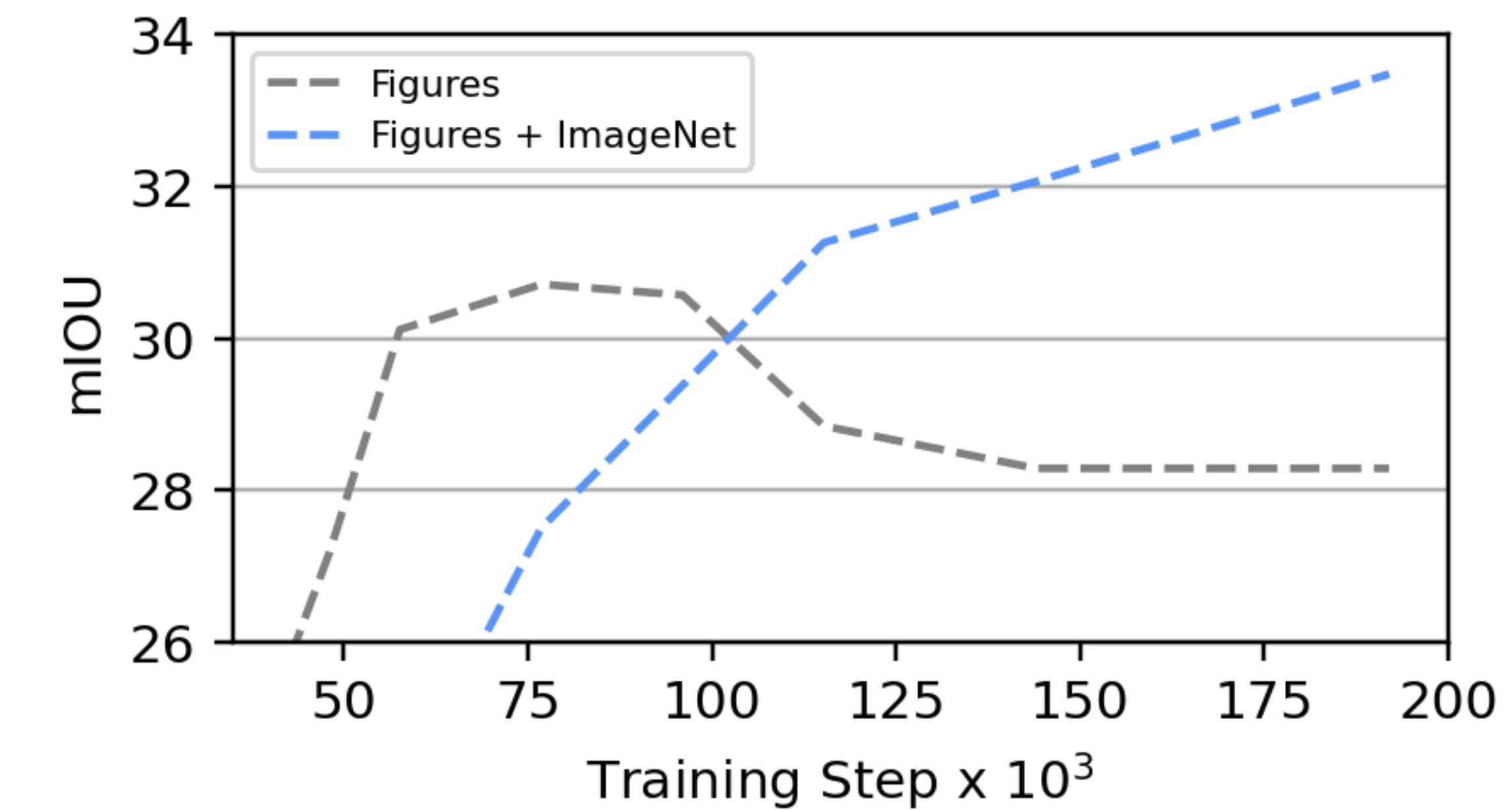
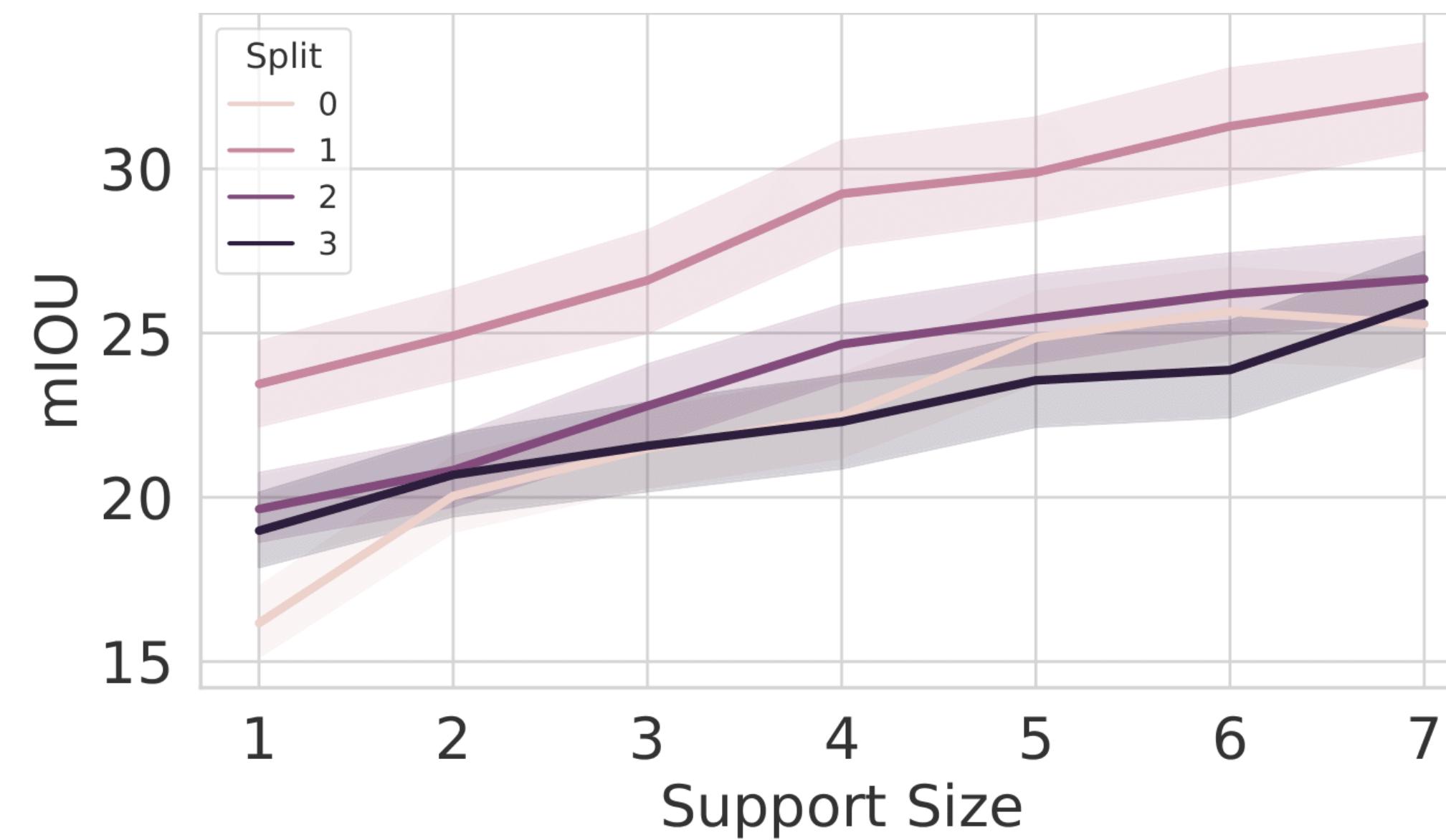
Prompt Variations:

Different prompts for the same task

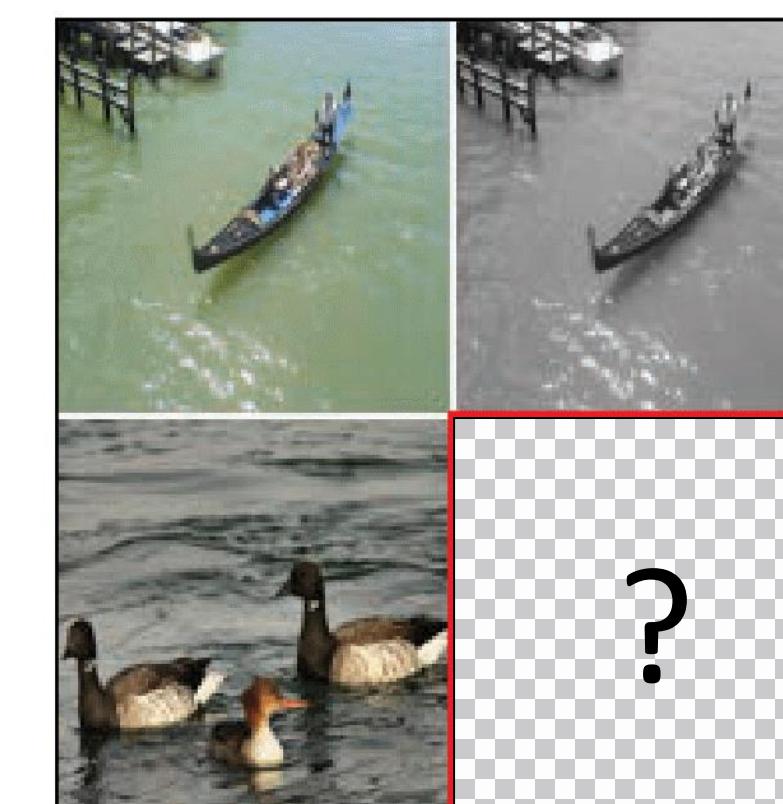
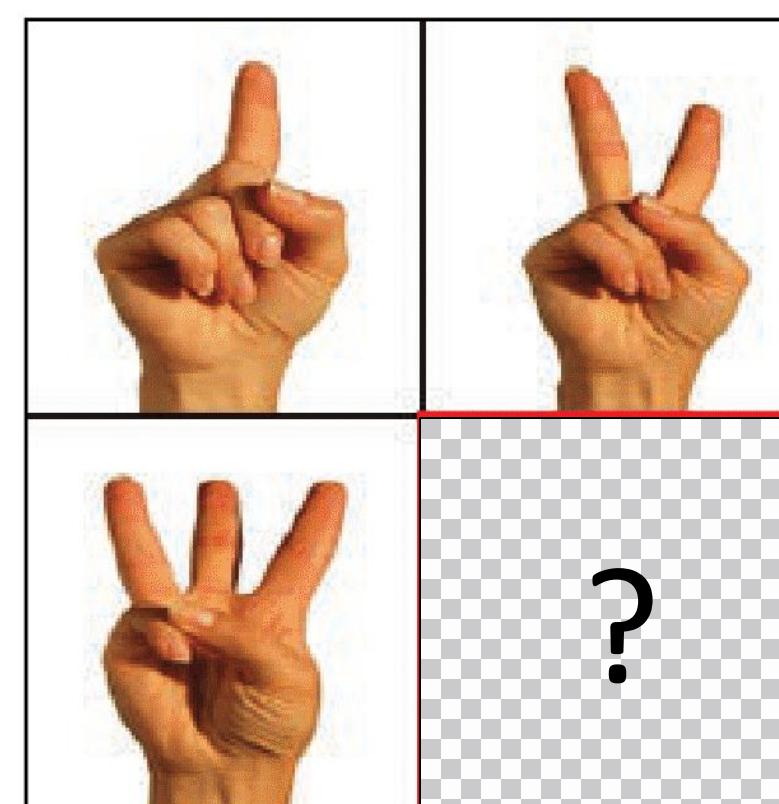
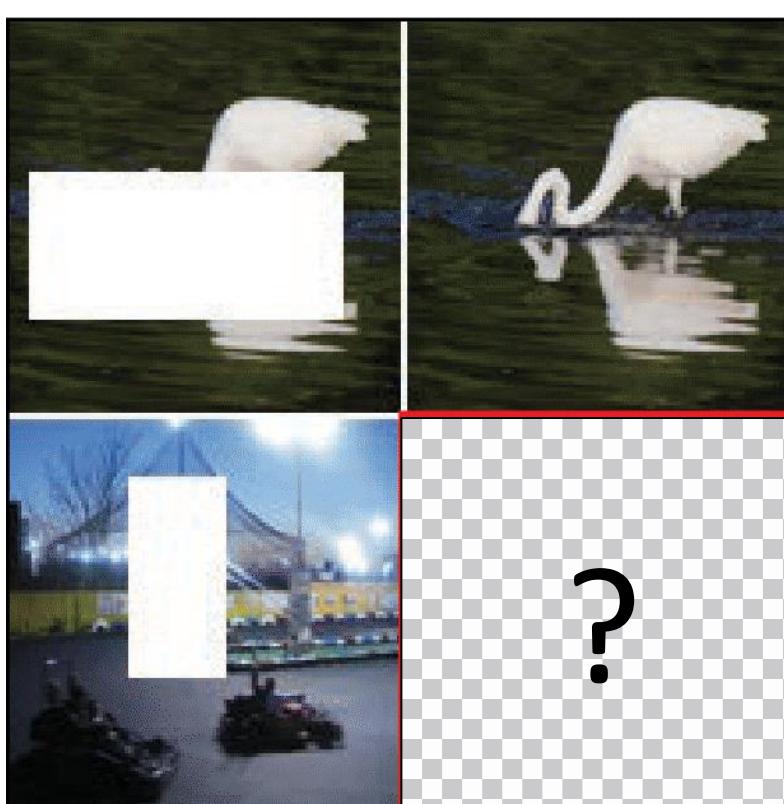
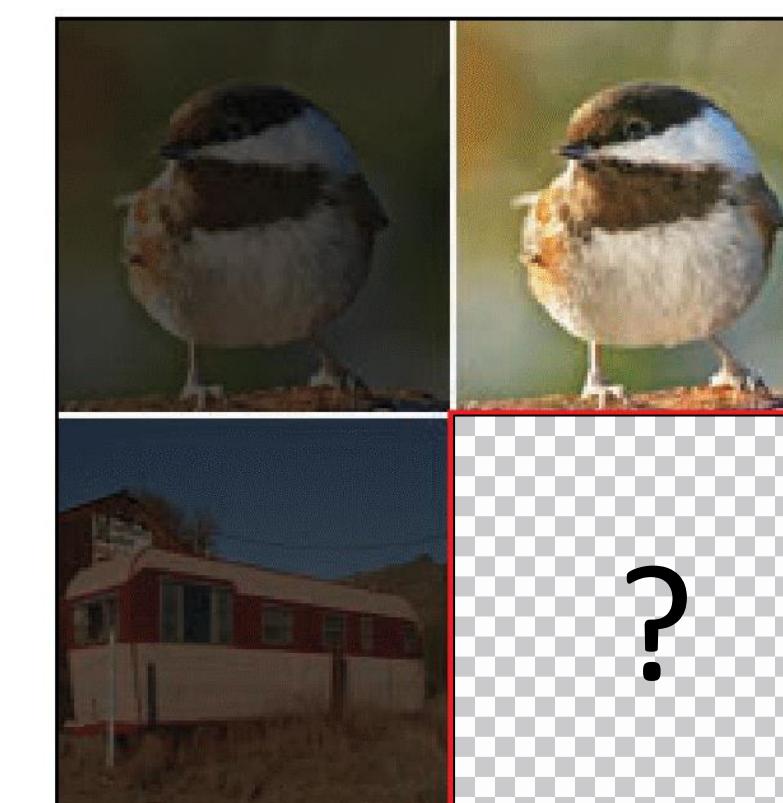
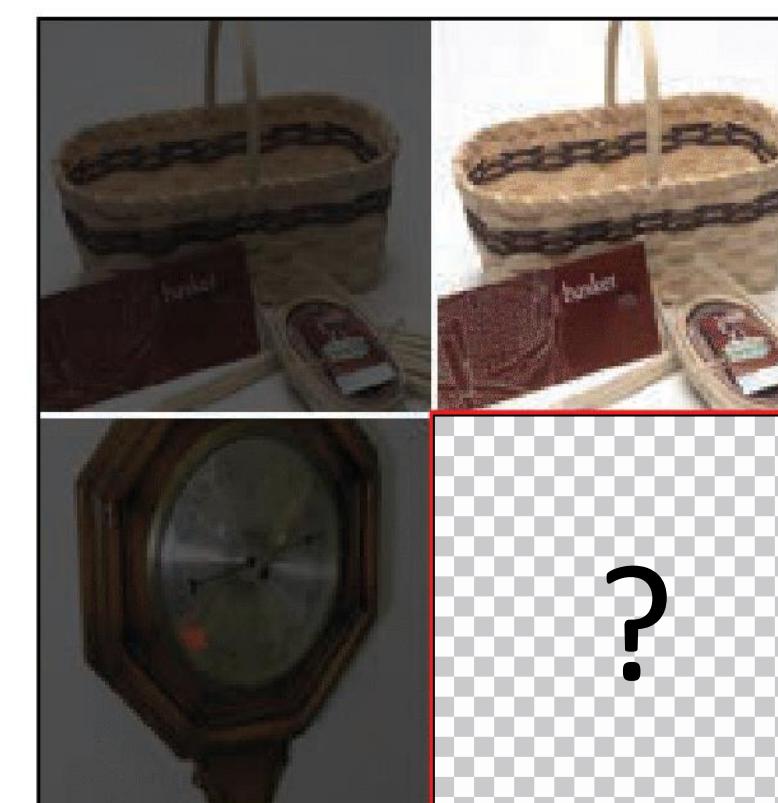
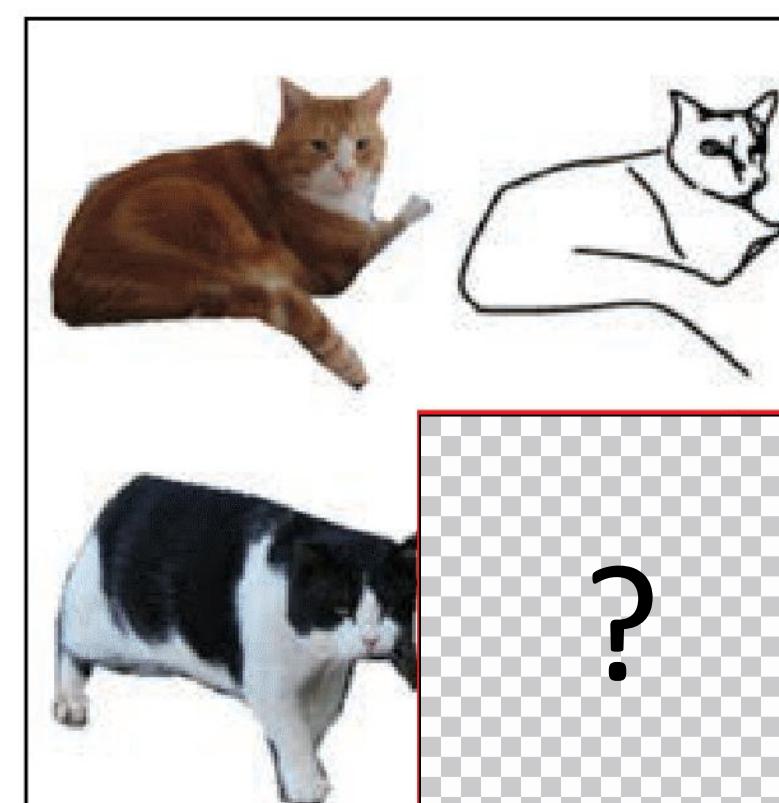
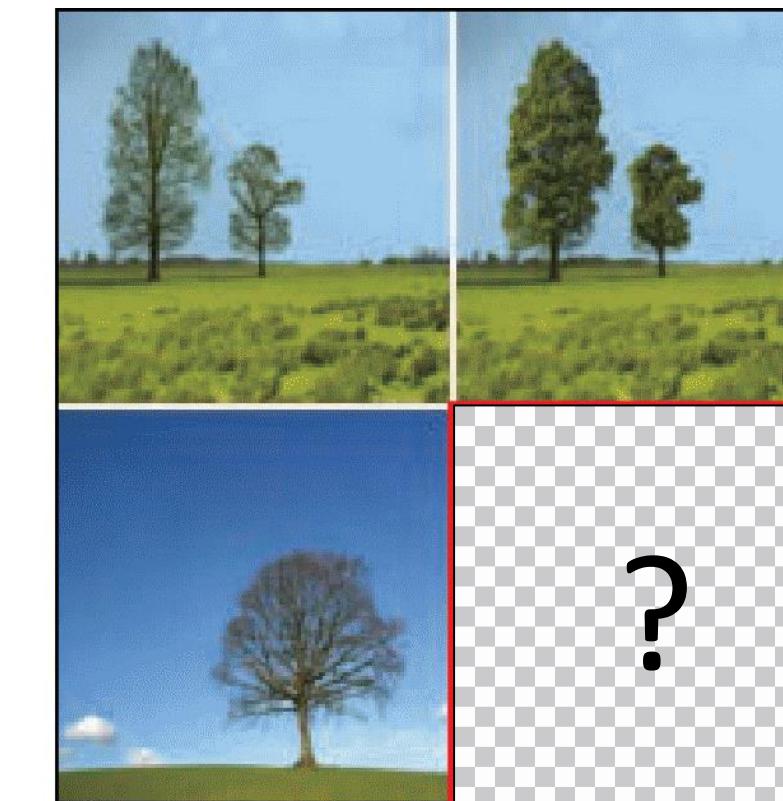
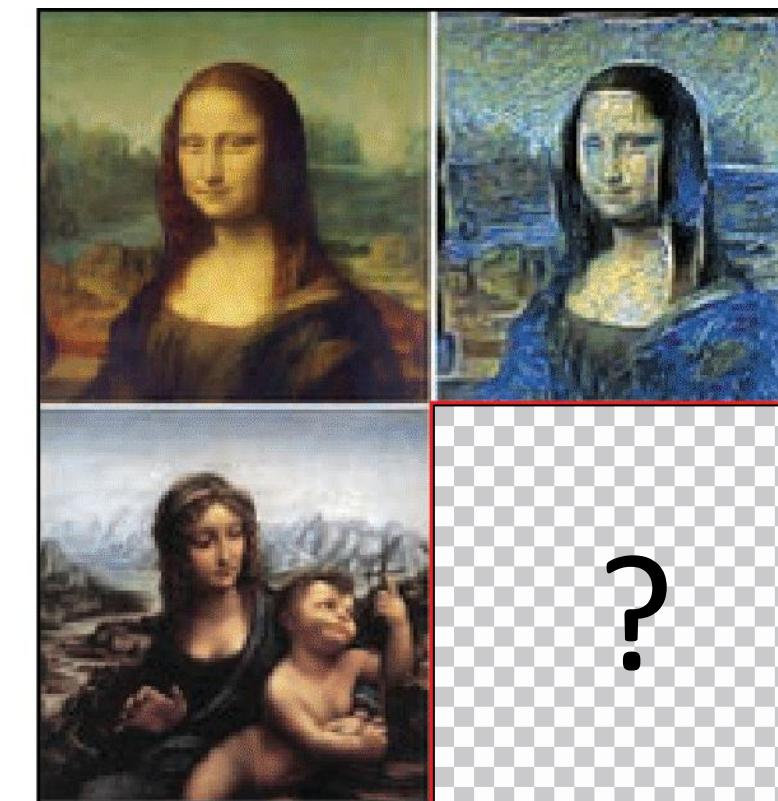
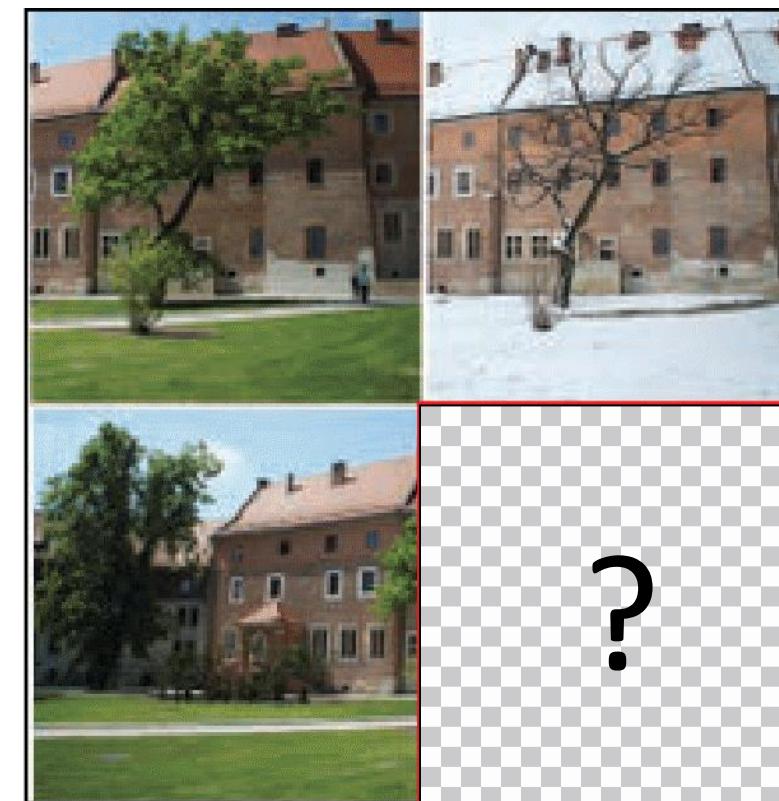
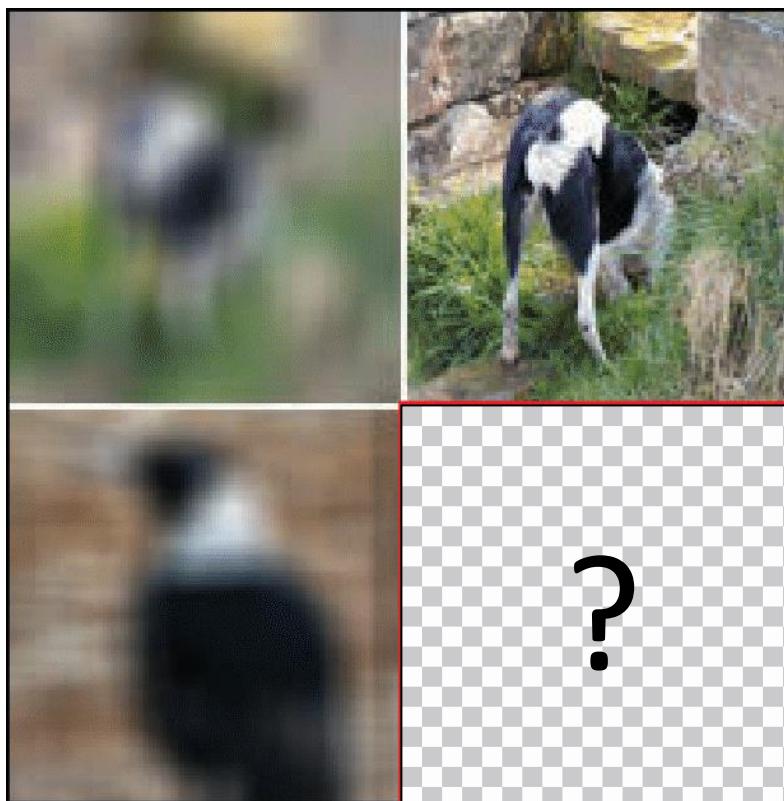


Ways to improve the performance

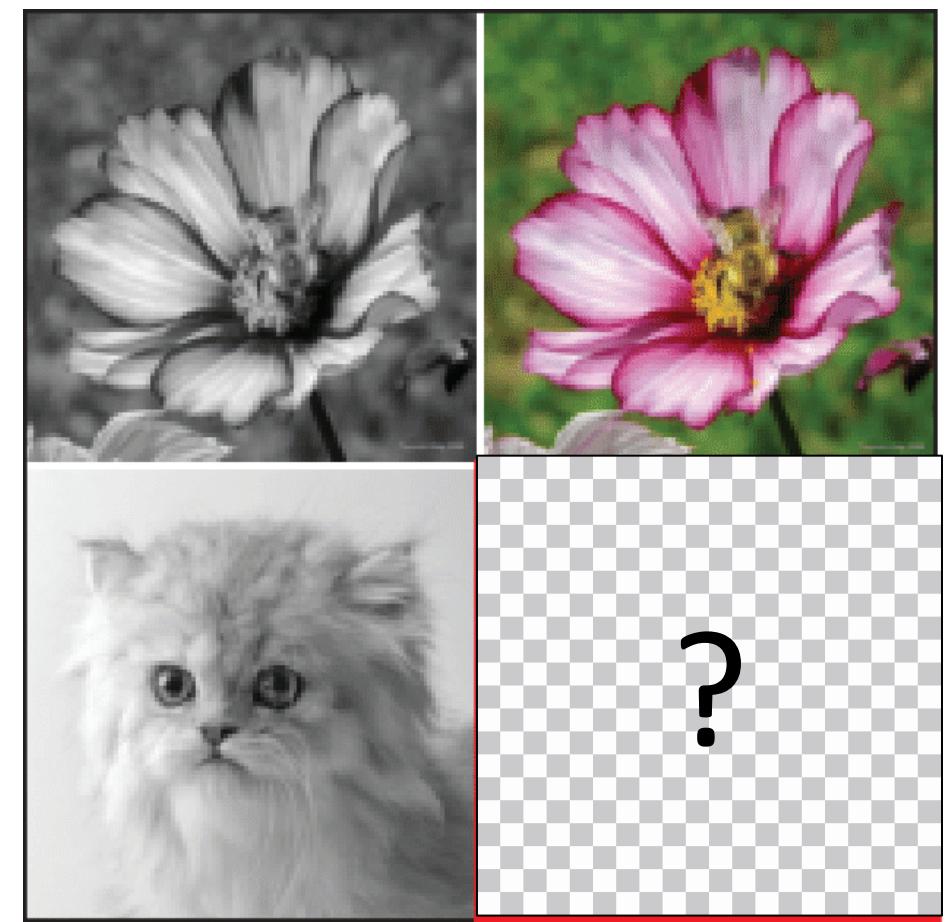
- More task examples
- Using more data (e.g, ImageNet)



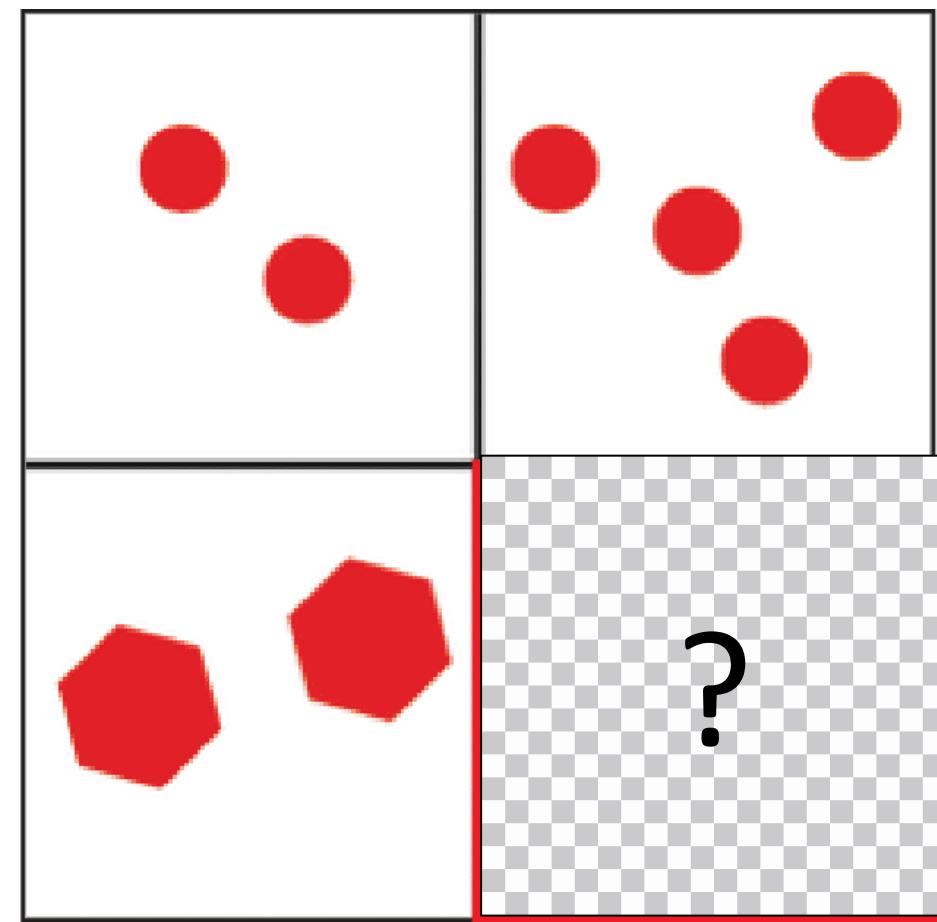
Various tasks



Limitations

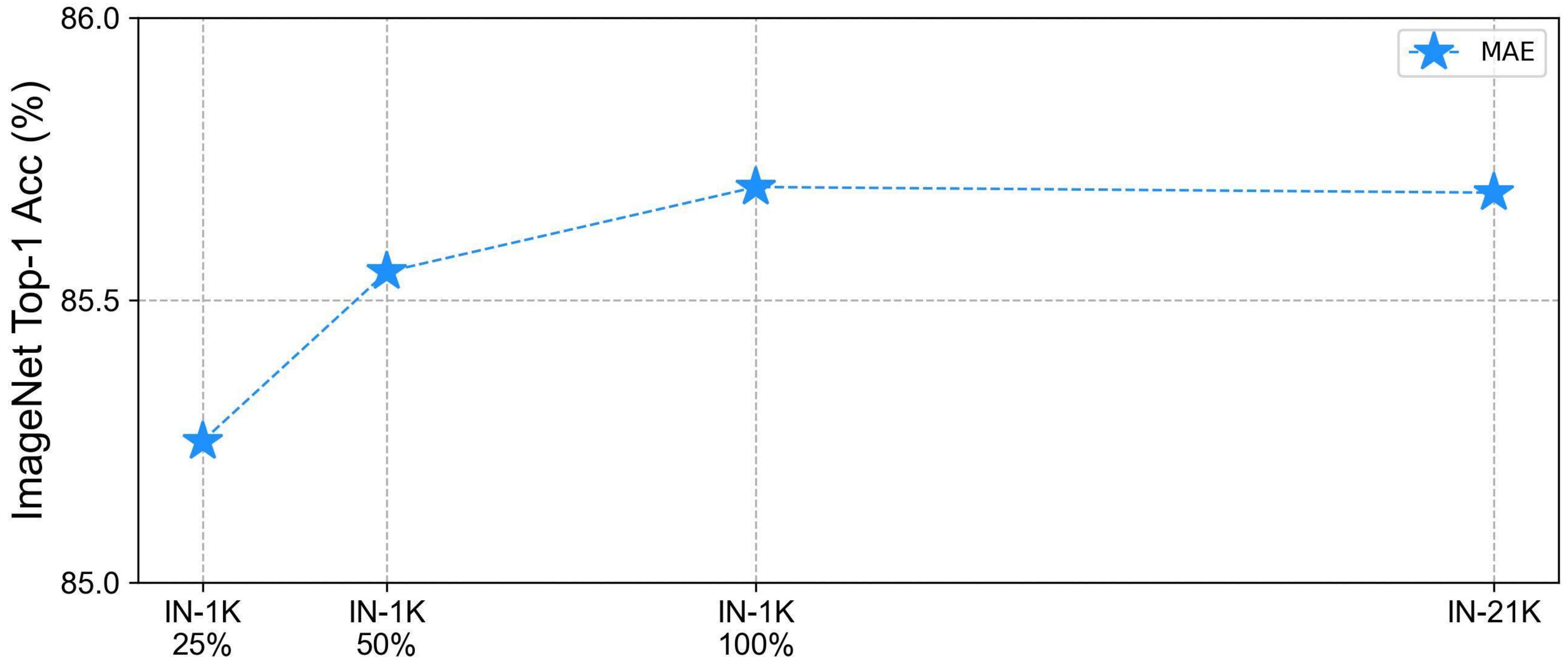


Task ambiguity



Non-aligned input-output

MAE doesn't seem to scale 😞



Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was still red

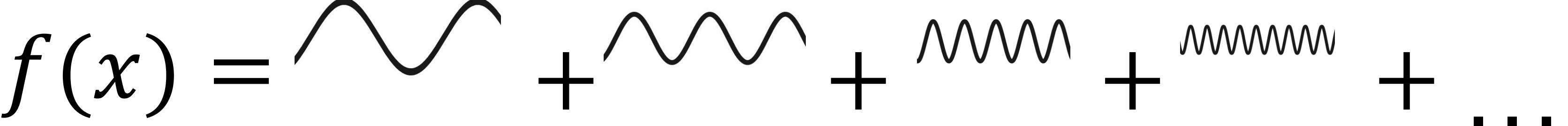
Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was still red

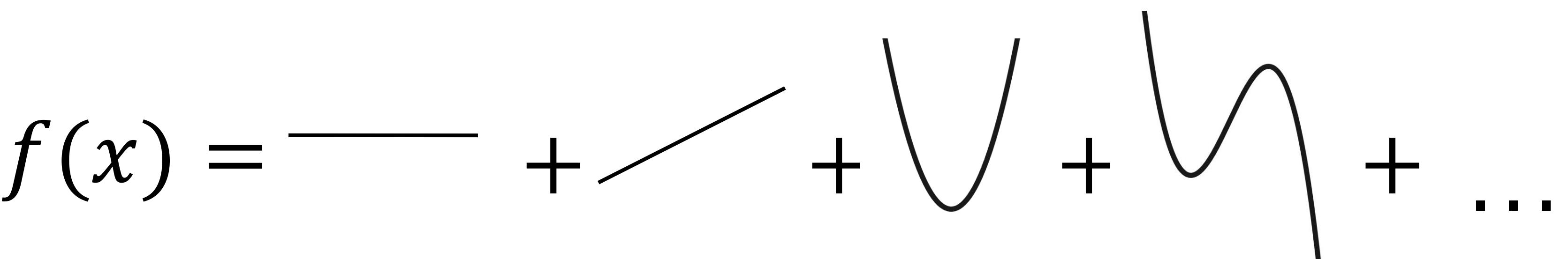
Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was still red

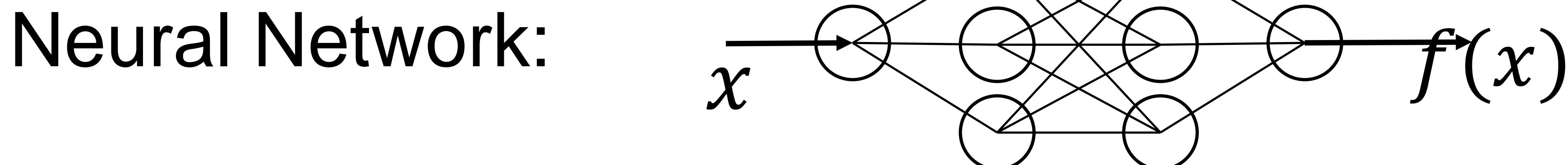
$$P(x_n | x_{n-1}, x_{n-2}, x_{n-3}, x_{n-4}, x_{n-5}, x_{n-6}, x_{n-7}, x_{n-8}, x_{n-9}, x_{n-10}, x_{n-11}, x_{n-12}, x_{n-13})$$

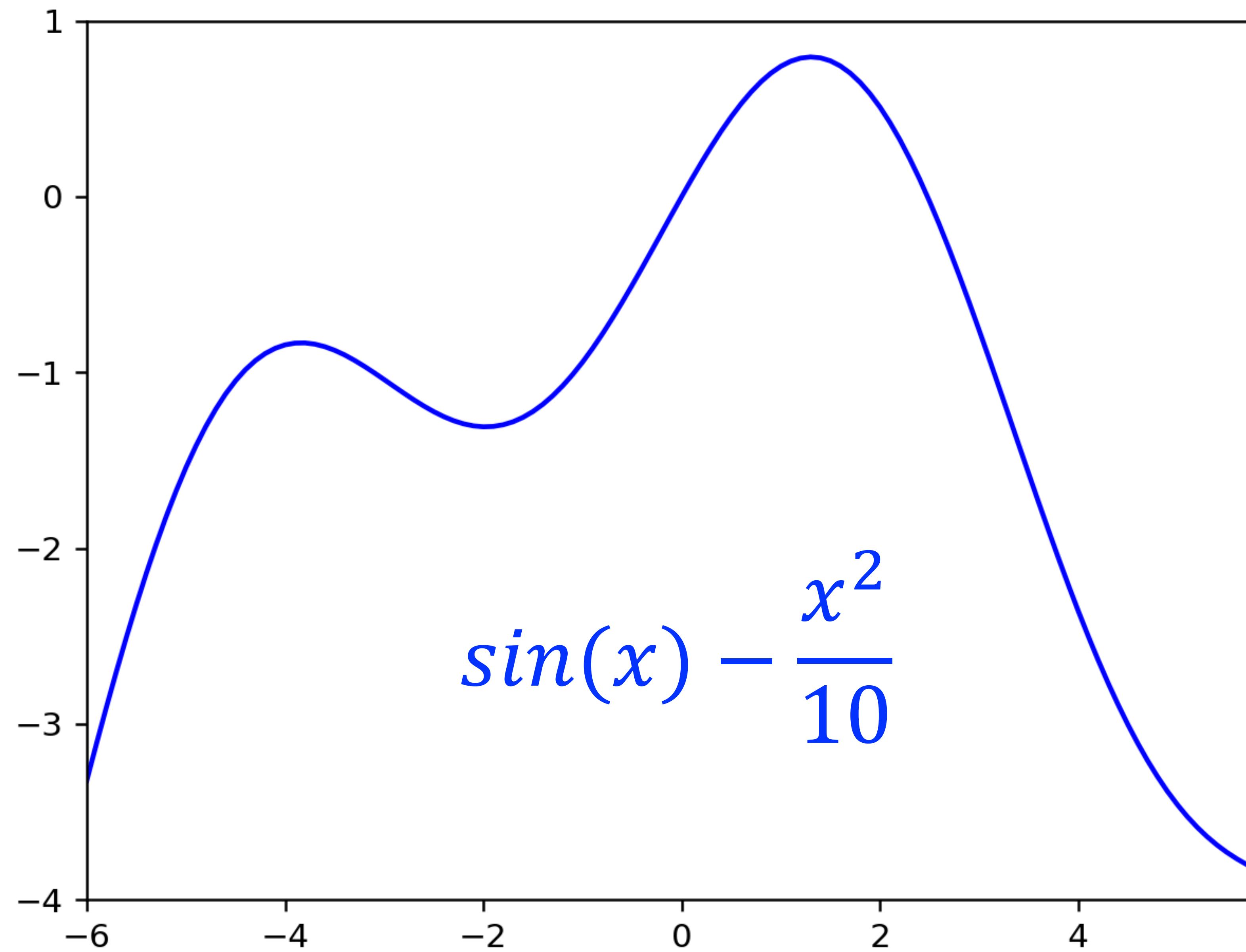
10^{70} combinations

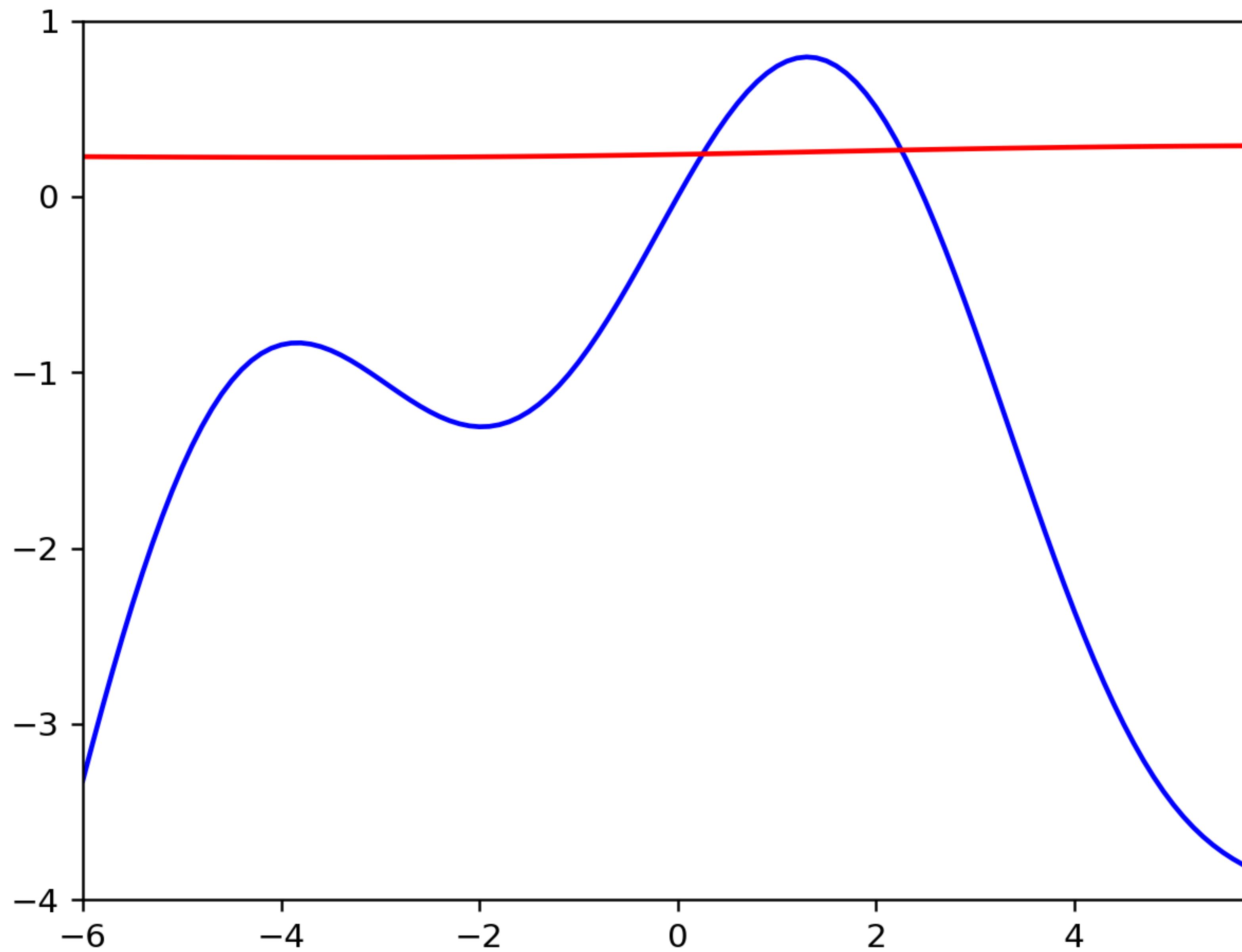
Function Approximation

Fourier Series: $f(x) =$ 

Taylor Series: $f(x) =$ 



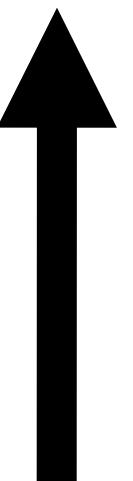




$$P(x_n | x_{n-1}, x_{n-2}, x_{n-3}, x_{n-4}, x_{n-5}, x_{n-6}, x_{n-7}, \dots)$$

Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was still red

red

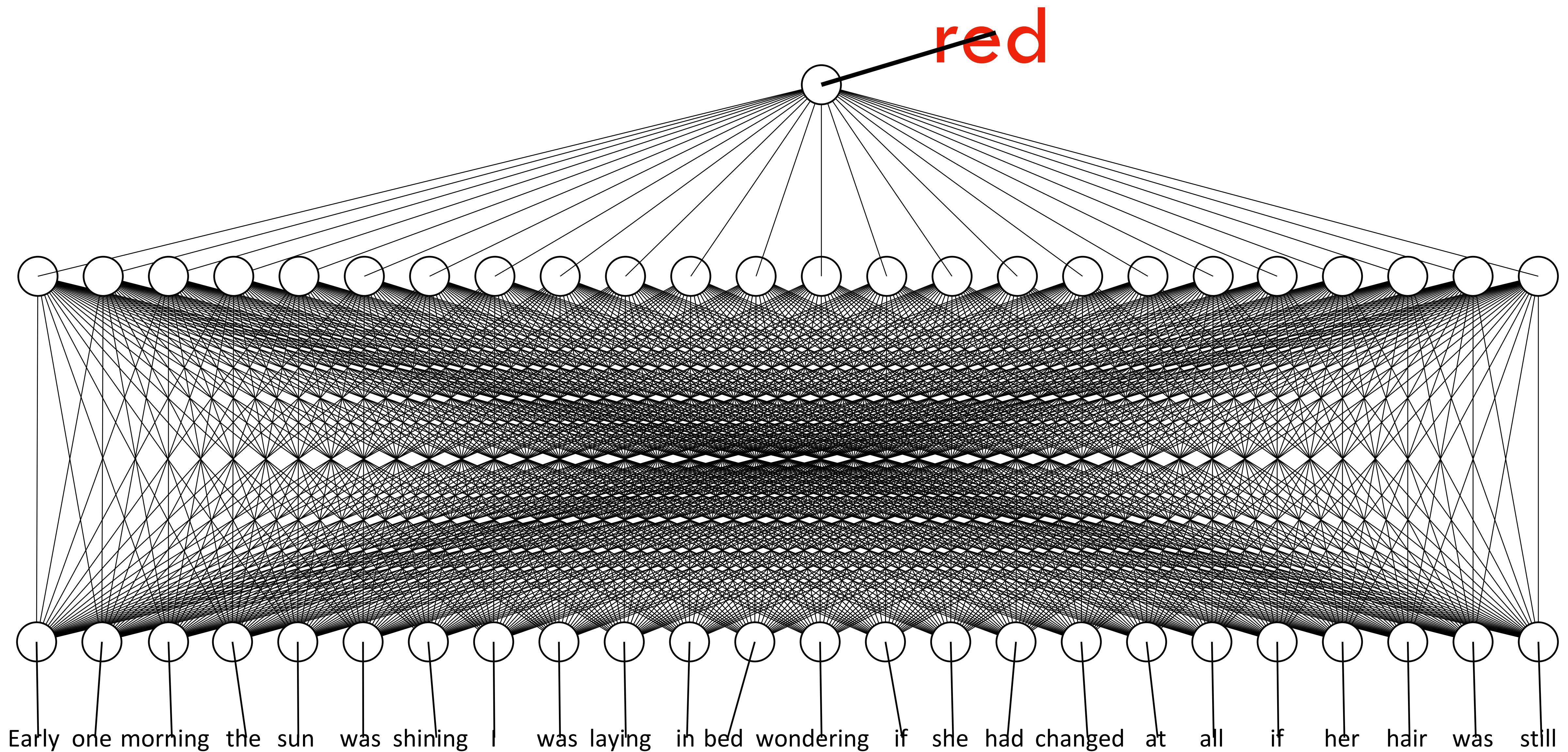


Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was still

red

neural network

Early one morning the sun was shining was laying in bed wondering if she had changed at all if her hair was still

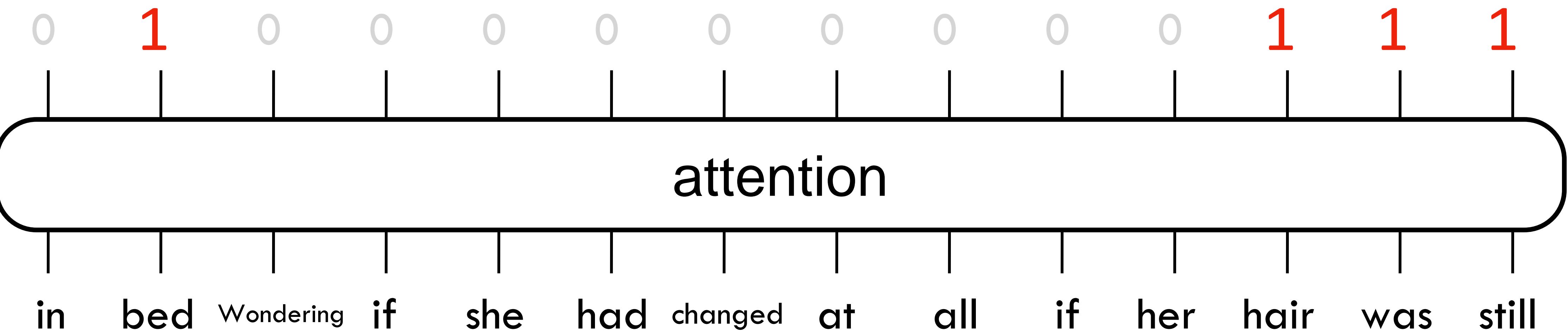


slide from Steve Seitz's [video](#)

Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was still ?

bed

hair was still red



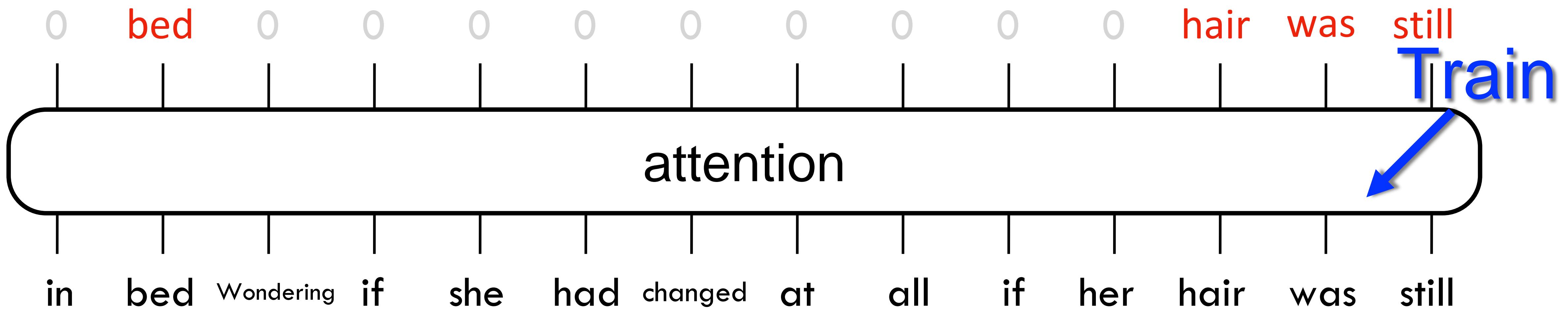
red

next word prediction

0 bed 0 0 0 0 0 0 0 hair was still

attention

in bed Wondering if she had changed at all if her hair was still



Train

red

next word prediction

0 bed 0 0 0 0 0 0 0 hair was still

attention

in bed Wondering if she had changed at all if her hair was still

Train

brown

next word prediction

0 bed 0 0 0 0 0 0 0 hair was still

attention

in bed Wondering if she had changed at all if her hair was still

Train

brown

next word prediction

0 bed 0 0 0 0 0 0 0 hair was still

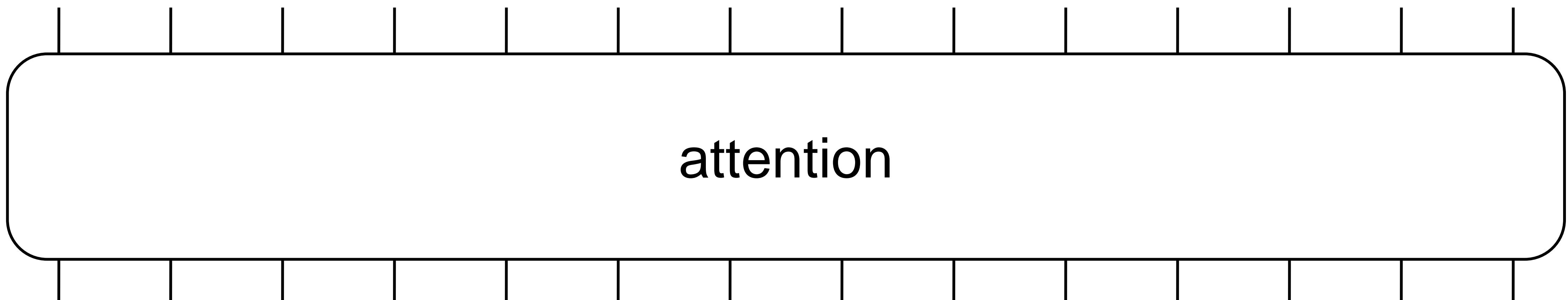
attention

in bed Wondering if she had changed at all if her hair was still

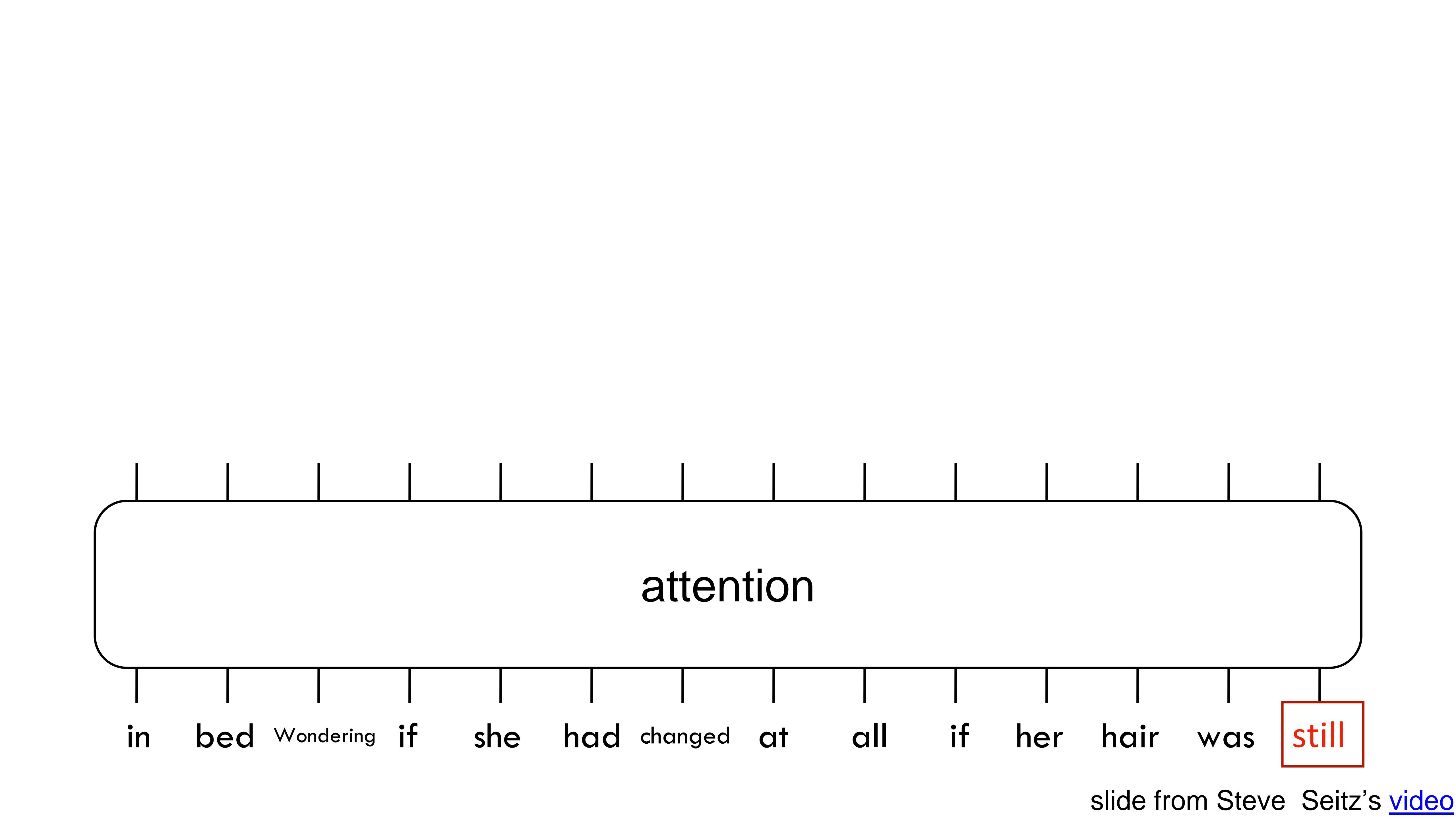
red

Transformer

in bed Wondering if she had changed at all if her hair was still

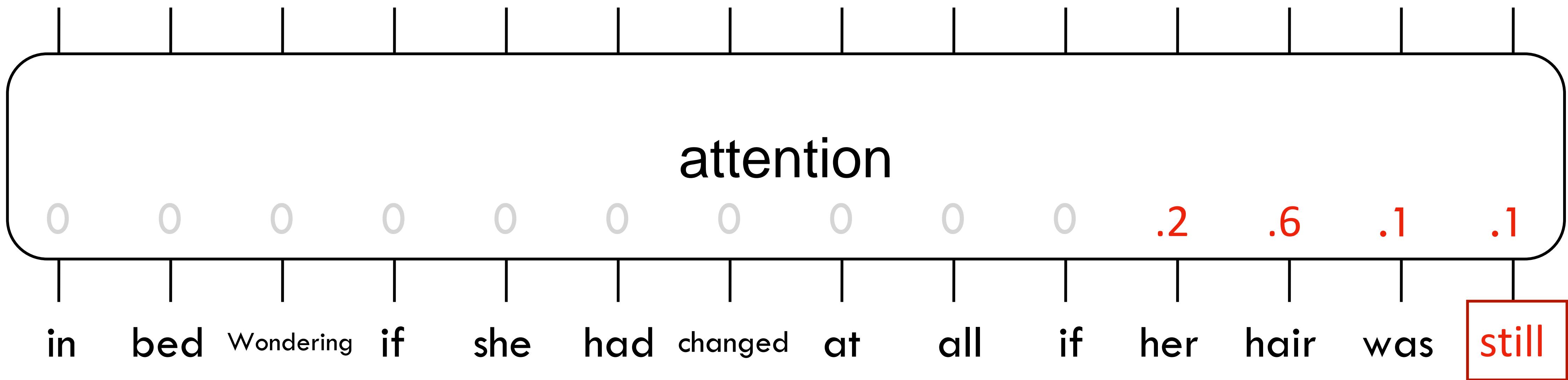


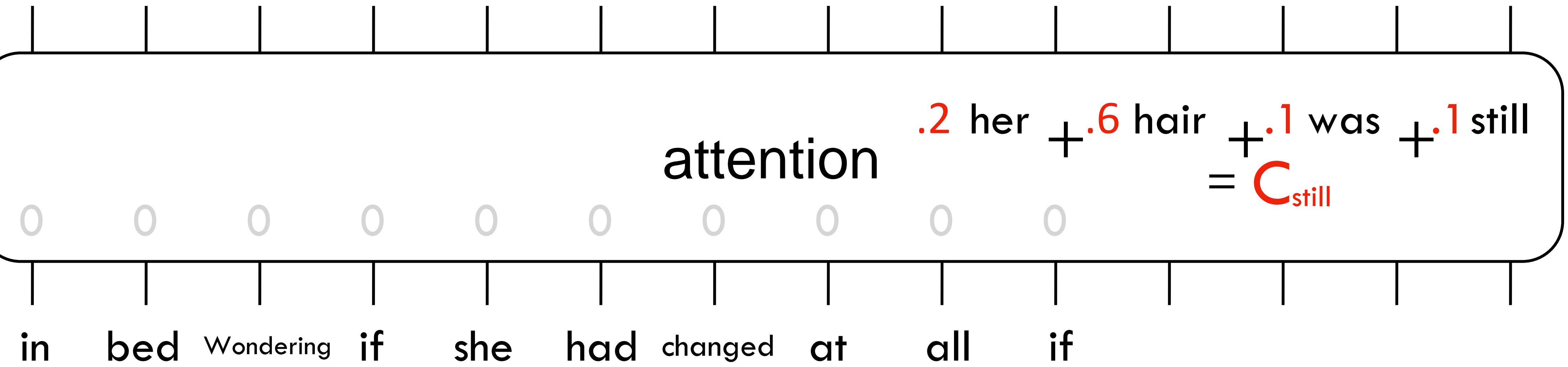
in bed Wondering if she had changed at all if her hair was still

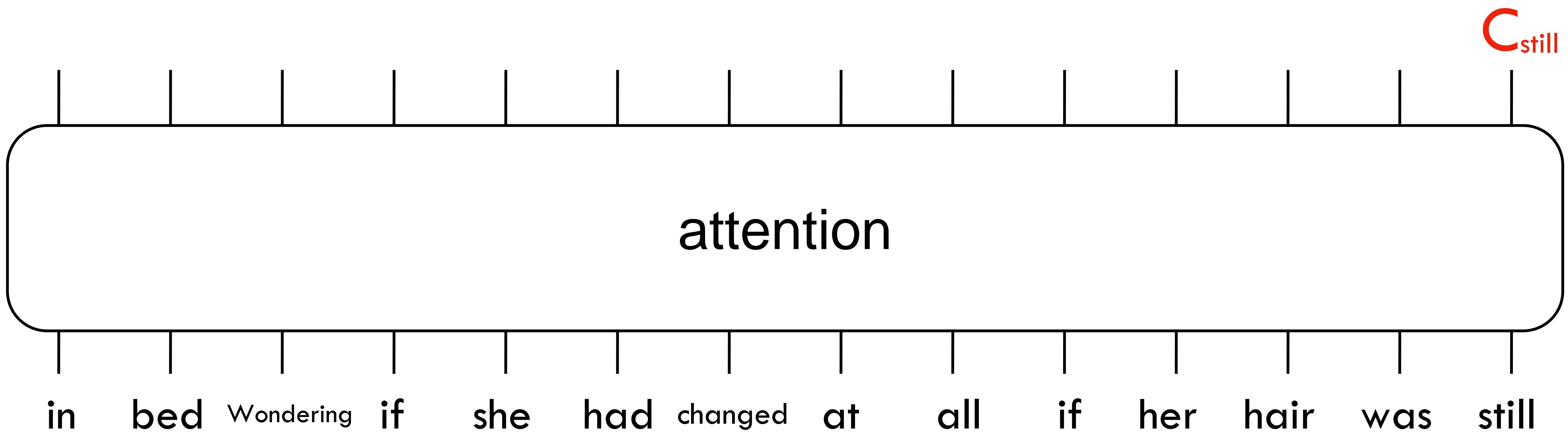


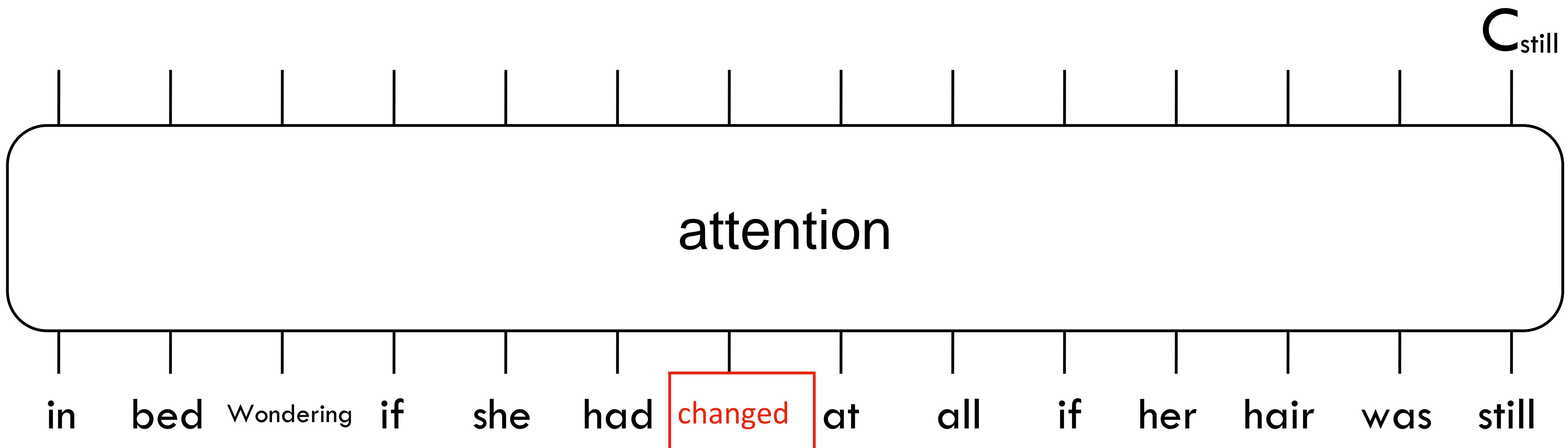
attention

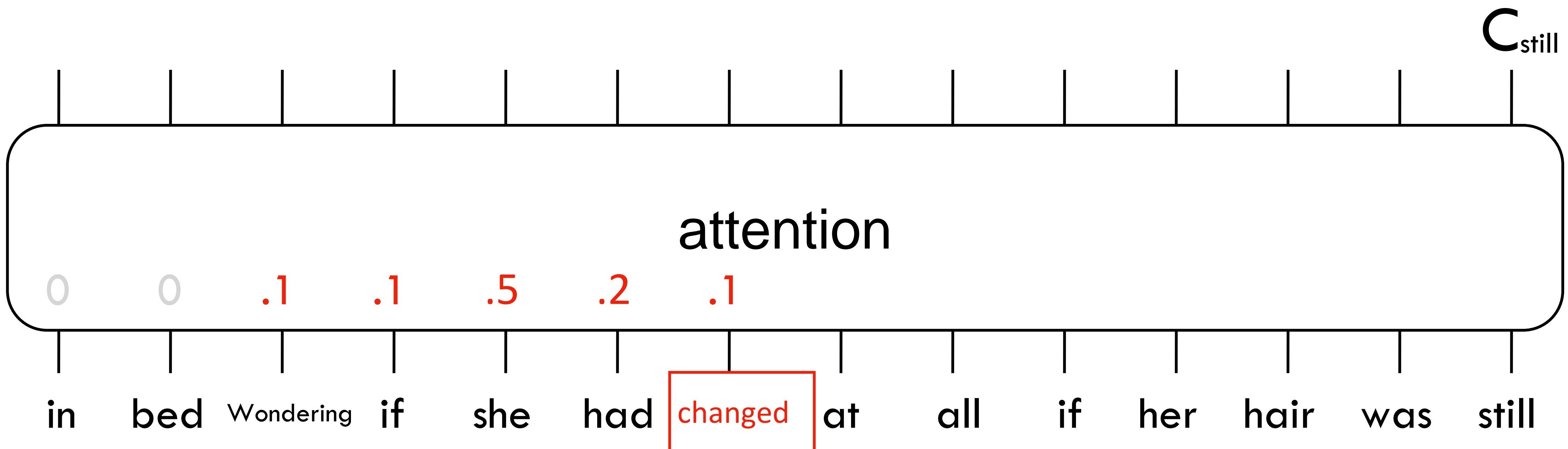
in bed wondering if she had changed at all if her hair was **still**

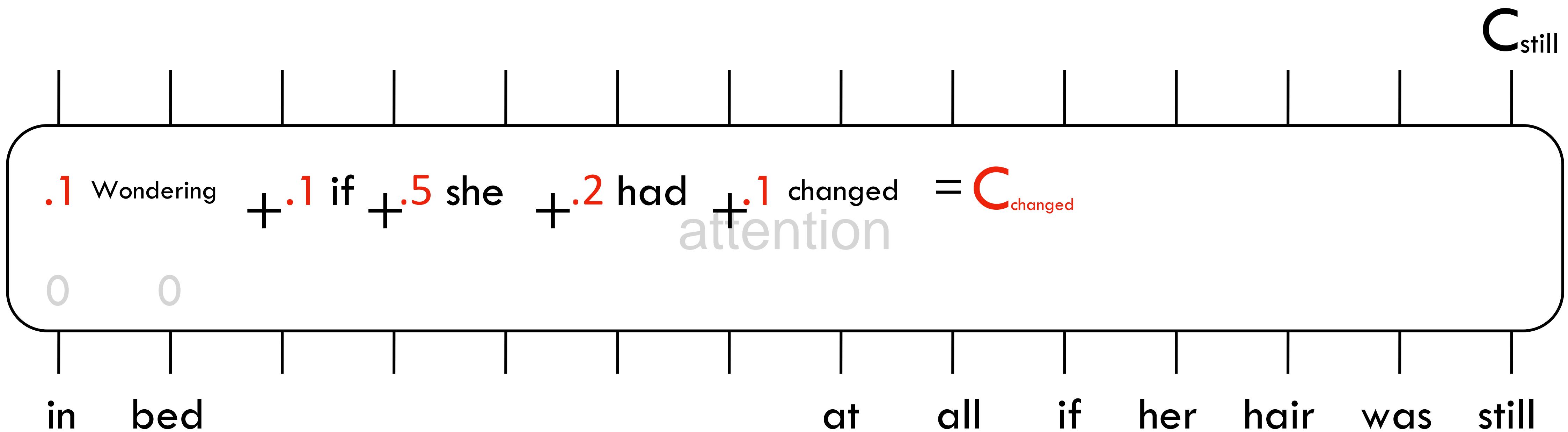


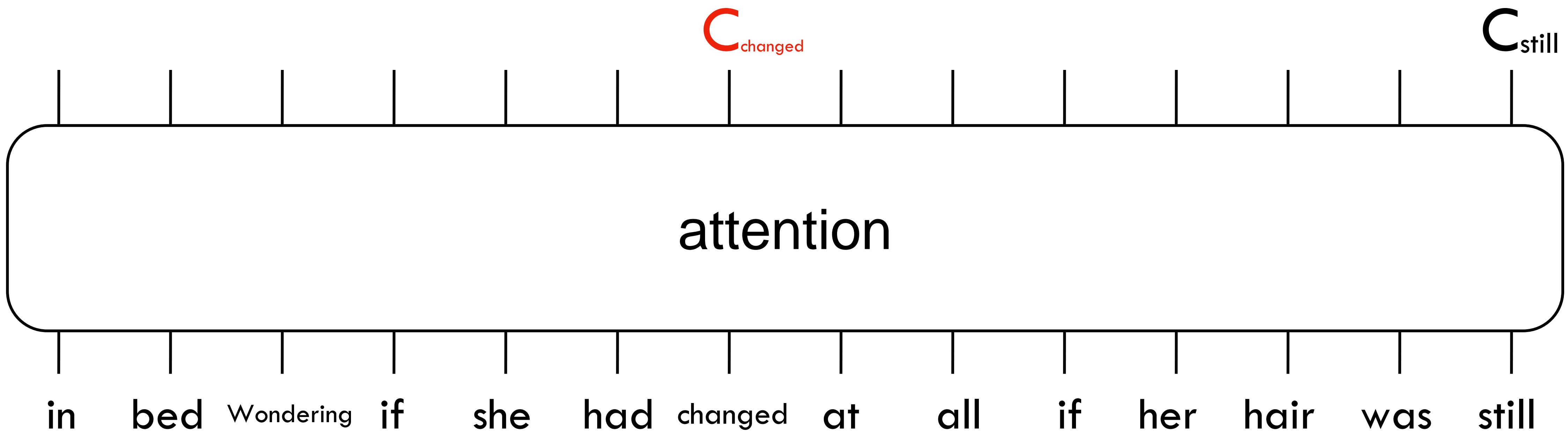












C_{in} C_{bed} $C_{wondering}$ C_{if} C_{she} C_{had} $C_{changed}$ C_{at} C_{all} C_{still} C_{still} C_{hair} C_{was} C_{still}

attention

in bed wondering if she had changed at all if her hair was still

prediction

C_{in} C_{bed} $C_{wondering}$ C_{if} C_{she} C_{had} $C_{changed}$ C_{at} C_{all} C_{still} C_{still} C_{hair} C_{was} C_{still}

attention

in bed wondering if she had changed at all if her hair was still

prediction

C_{in} C_{bed} $C_{wondering}$ C_{if} C_{she} C_{had} $C_{changed}$ C_{at} C_{all} C_{still} C_{still} C_{hair} C_{was} C_{still}
in bed wondering if she had changed at all if her hair was still

attention

a

prediction

attention

It's

a the looking possible getting
0.4 0.3 0.1 0.1 0.1

a

prediction

attention

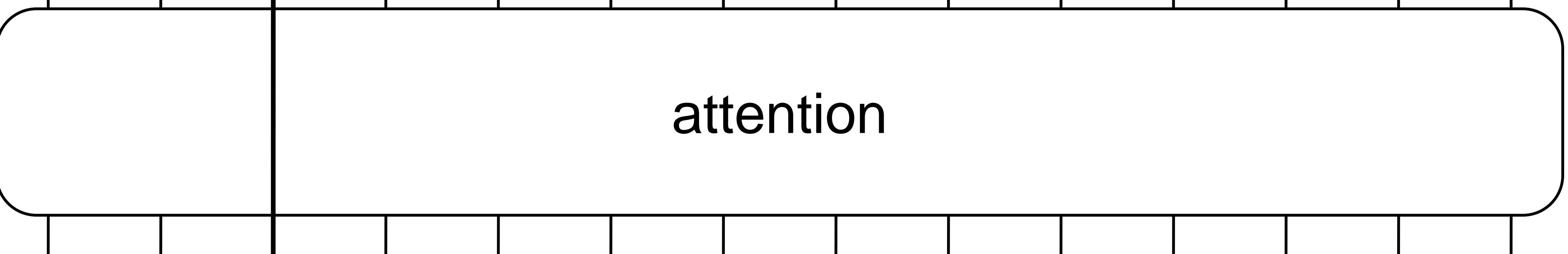
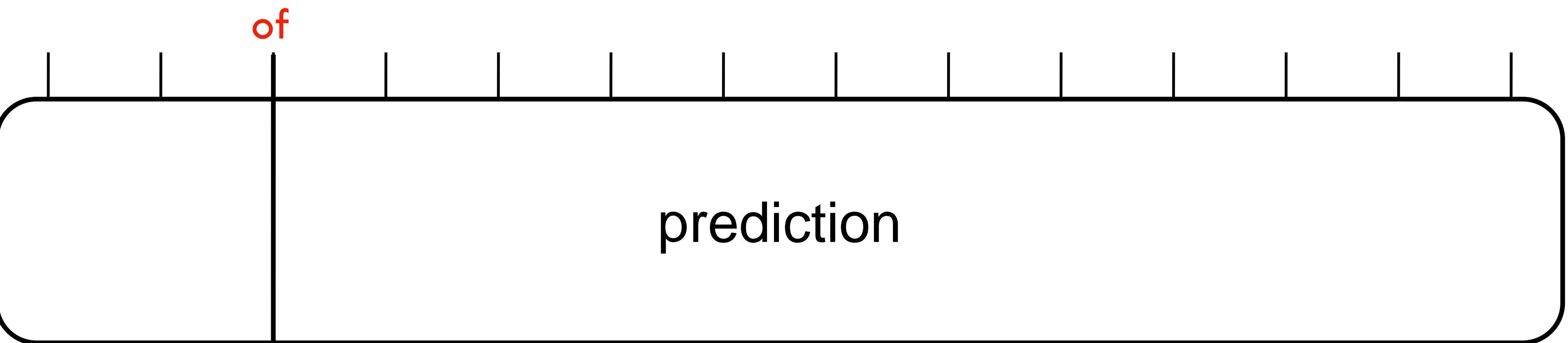
It's

lot

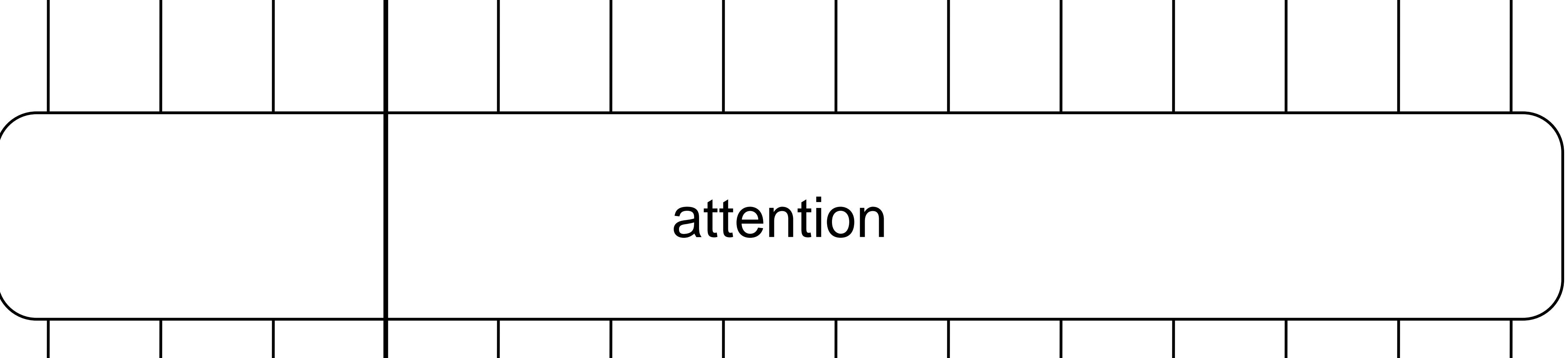
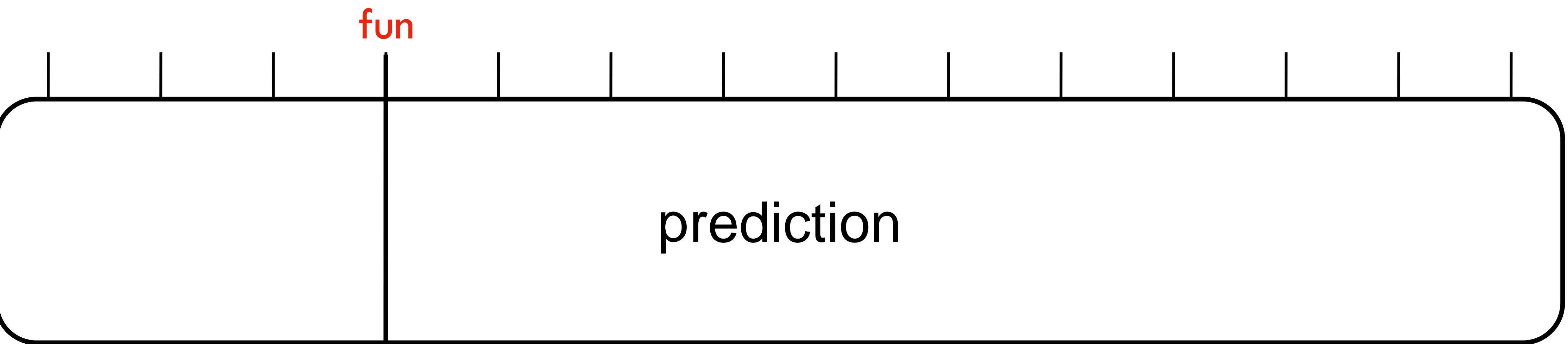
prediction

attention

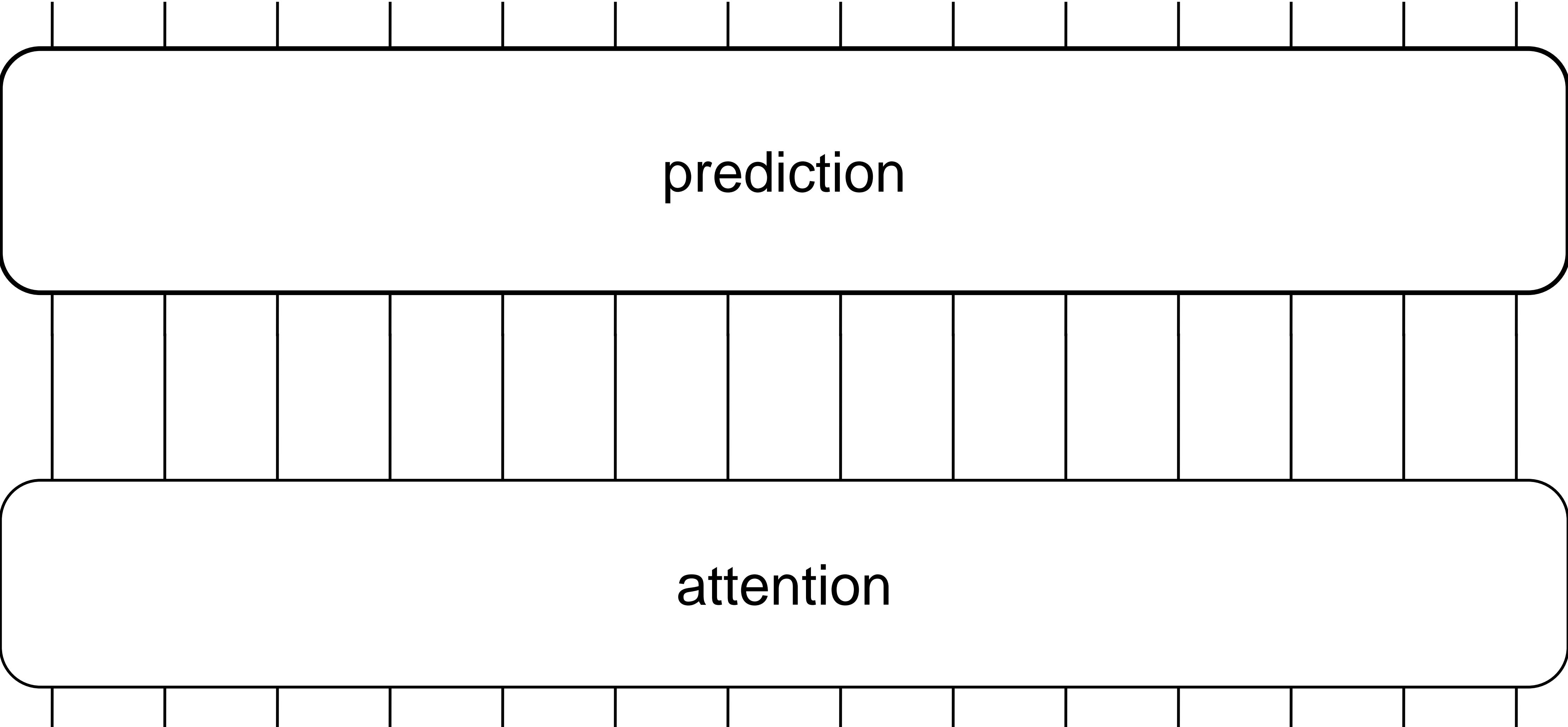
It's a



It's
a
lot



It's a lot of



prediction

attention

It's a lot of fun

Abraham

prediction

attention

The 16th was
president

The 16th President was ?

The capital of Zimbabwe is ?

Frank Zappa's middle name is ?

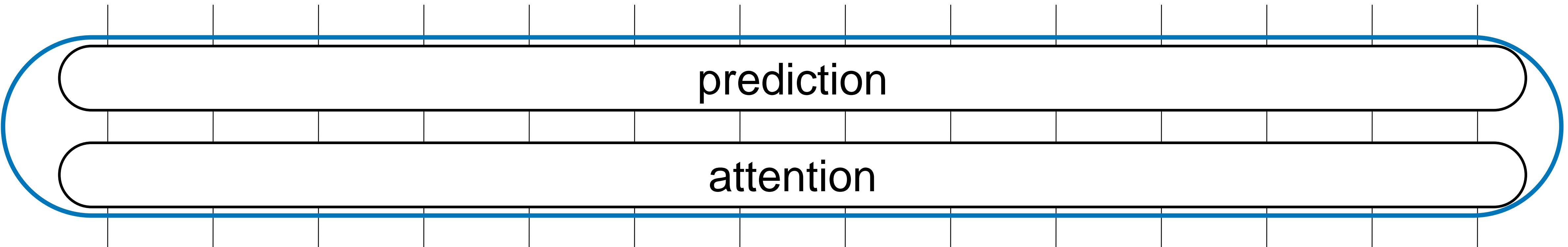
Napoleon was born on this date ?

The prime factorization of 19456721434 is ?

Queen Victoria's maiden name was ?

US per-capita income in 1957 was ?

The lat long coordinates of Rome are ?

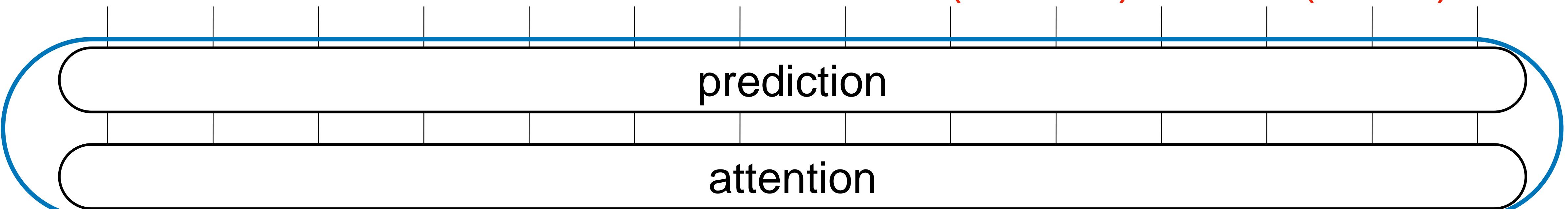


prediction

attention

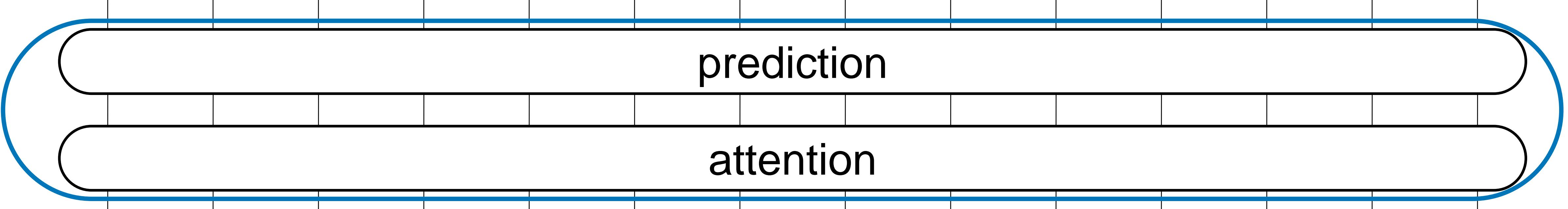
⋮

96 (GPT-3) 118 (Palm)



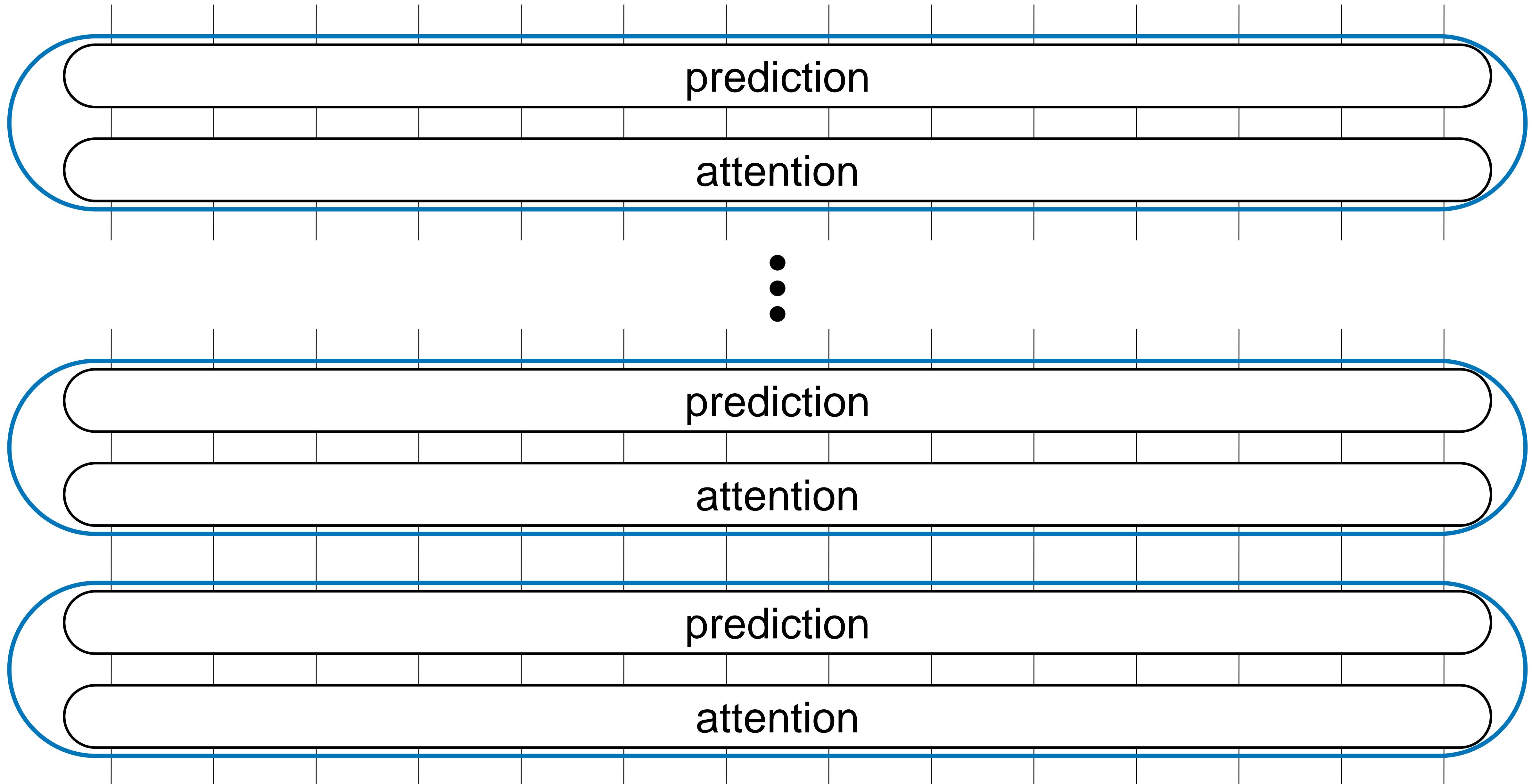
prediction

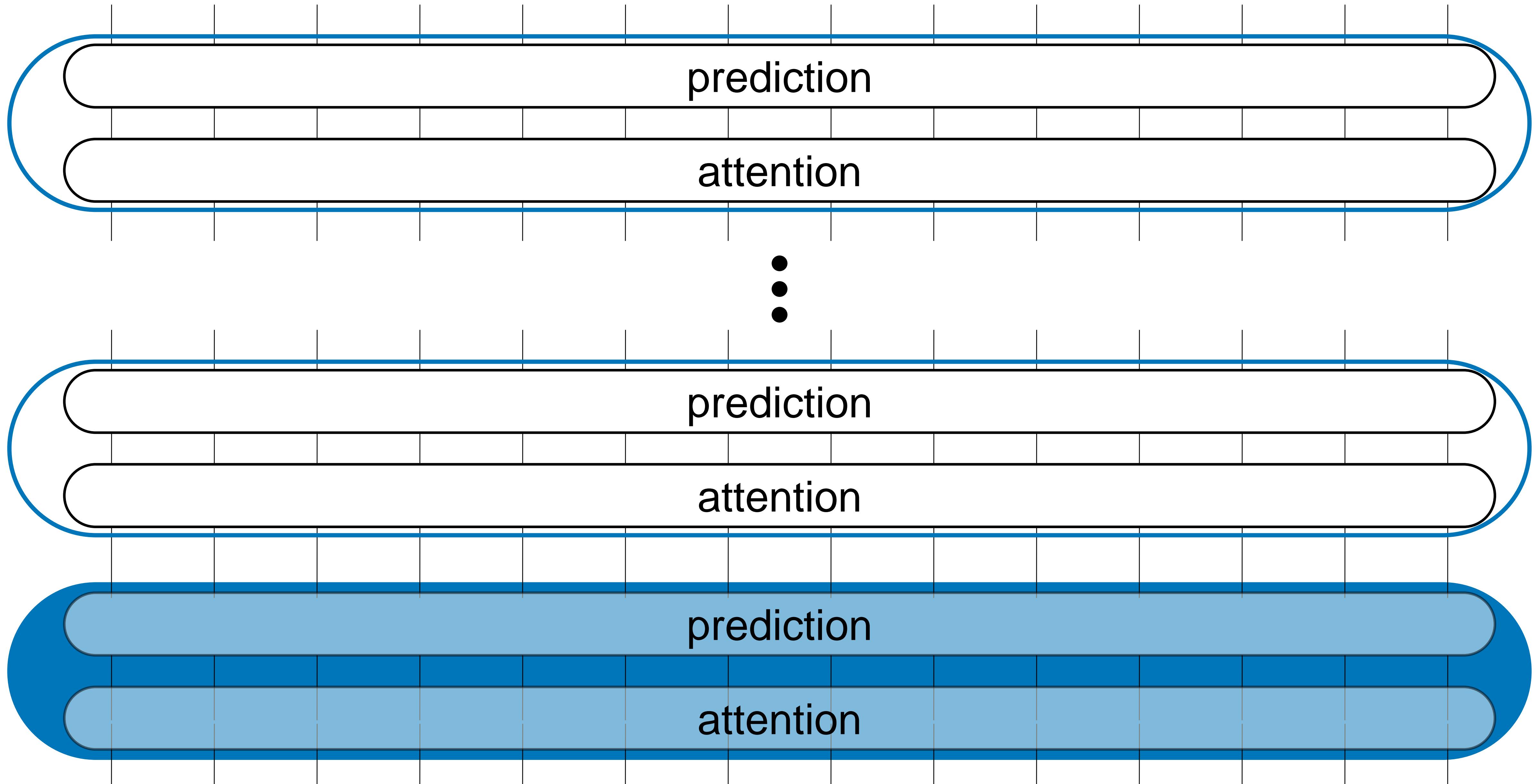
attention



prediction

attention

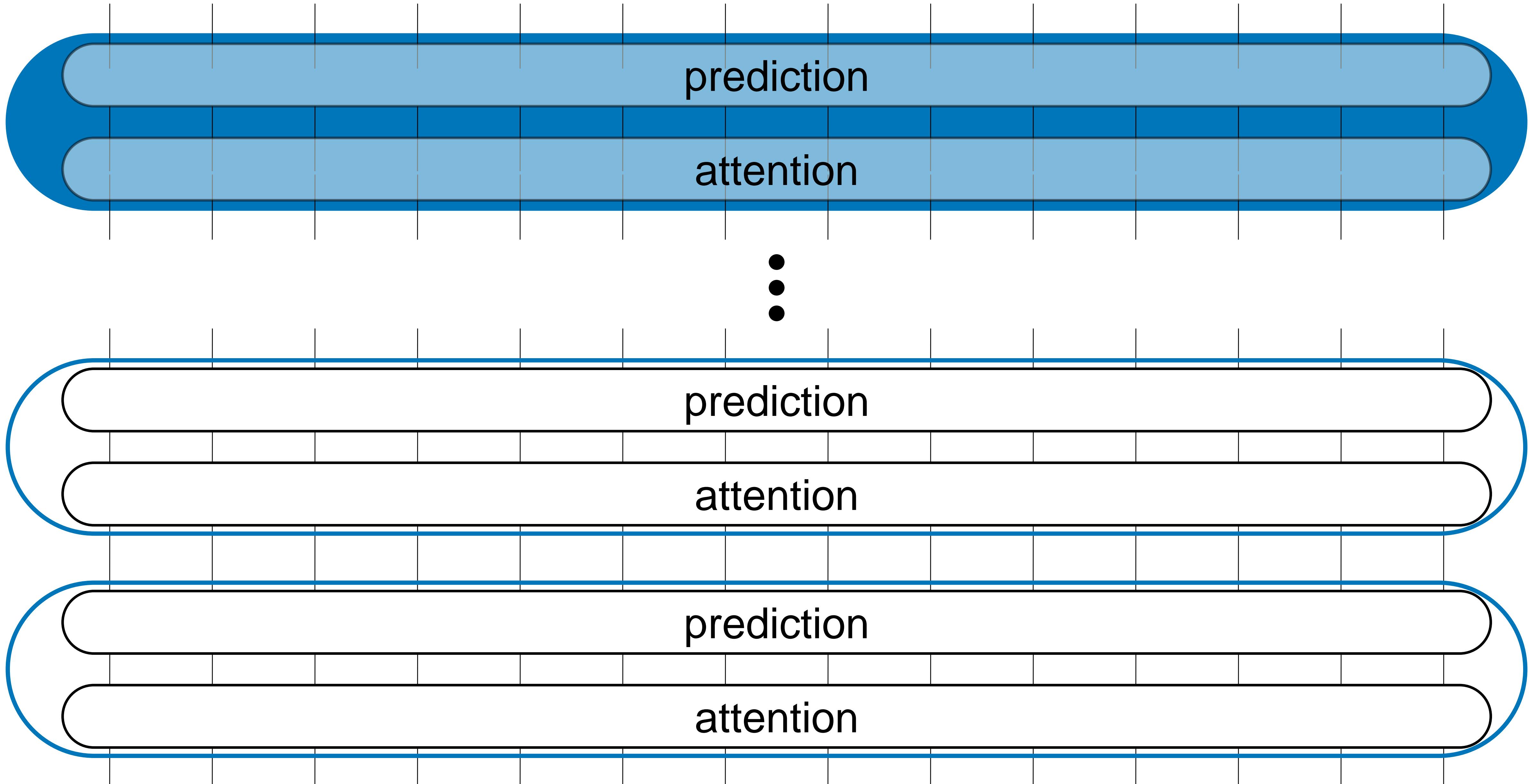


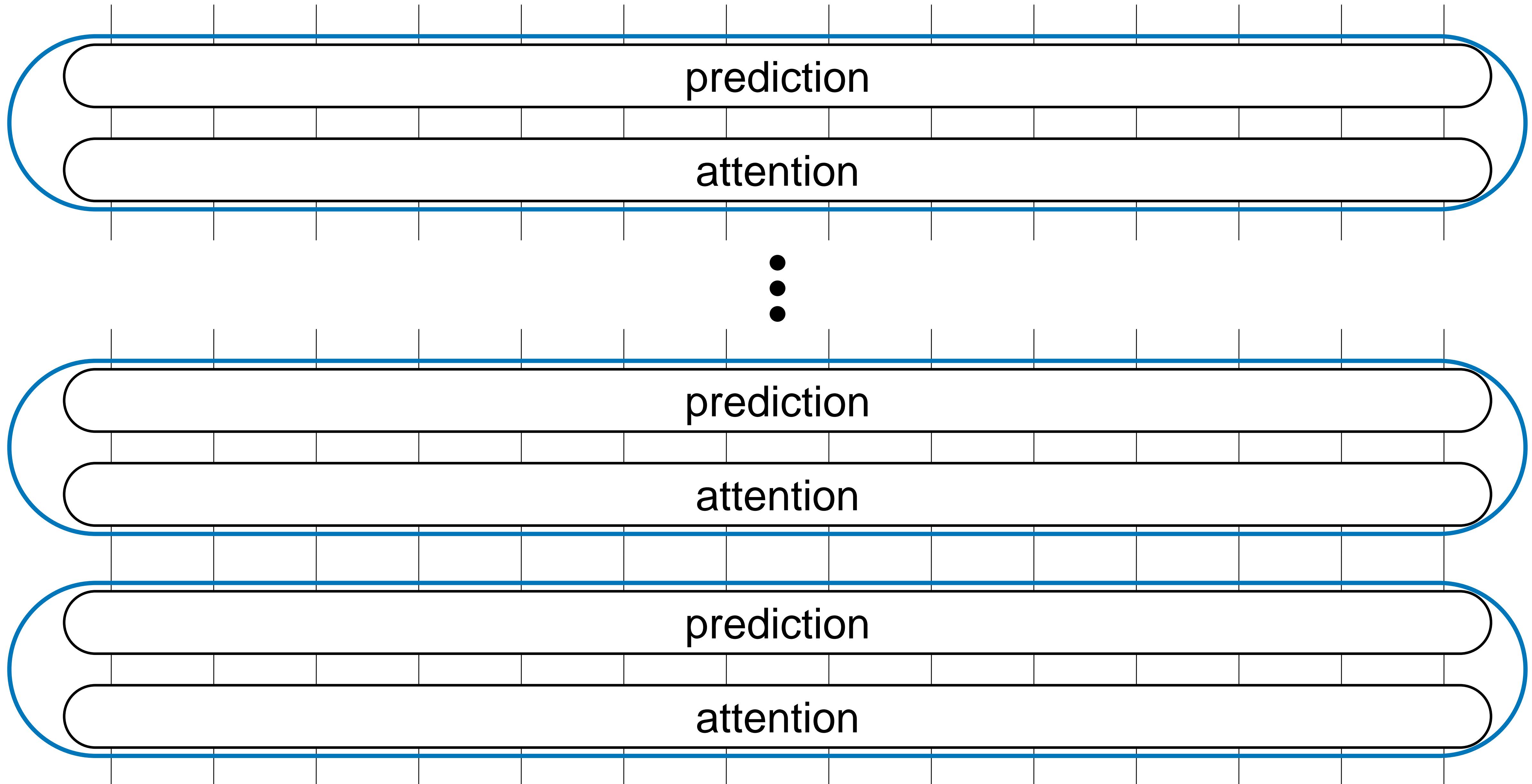


Syntax

slide from Steve Seitz's [video](#)

Semantics





How much data
to train?

All of it...
.

a pattern of characters that looks like a star

```
•   O   •  
○ ○ ○  
•   O   •
```

a pattern of characters that looks like a vertical line

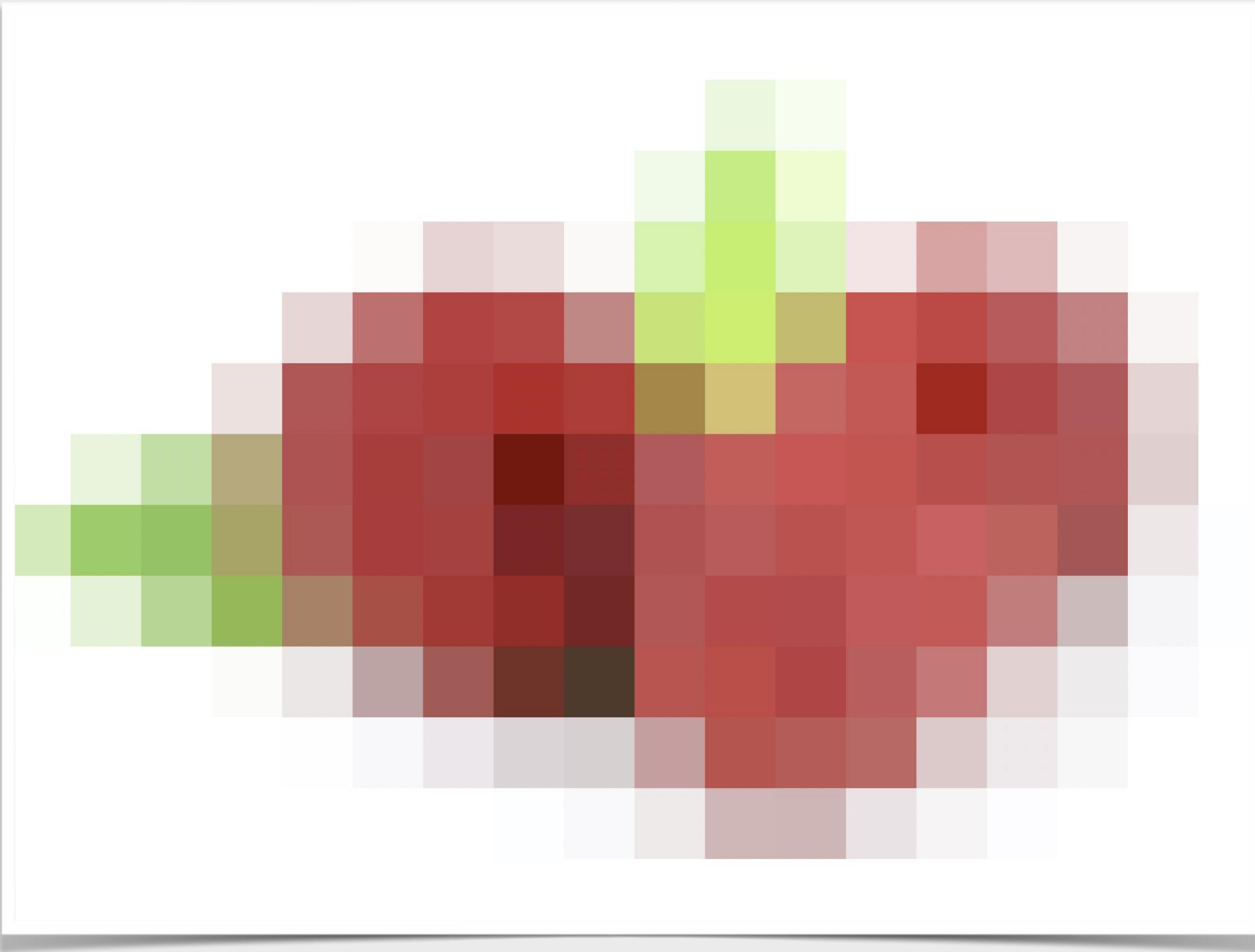
- O •
- O •
- O •
- O •
- O •

a pattern of characters that looks like a triangle

- ○ •
- ○ ○ •
- ○ ○ ○ •
- ○ ○ ○ ○ •
- ○ ○ ○ ○ ○ •



slide from Steve Seitz's [video](#)



slide from Steve Seitz's [video](#)



white white white white white white white white white green white white white white white
white white white white white white white white green green green white red red white
white white white white red red red red green green brown red red red red
white white white red red red red red brown green red red red red red
white green brown red
green green brown red
white green green brown red white
white white white white red red red black red red red red red white white
white white white white white white white white red red red red white white white
white white white white white white white white white red red white white white white

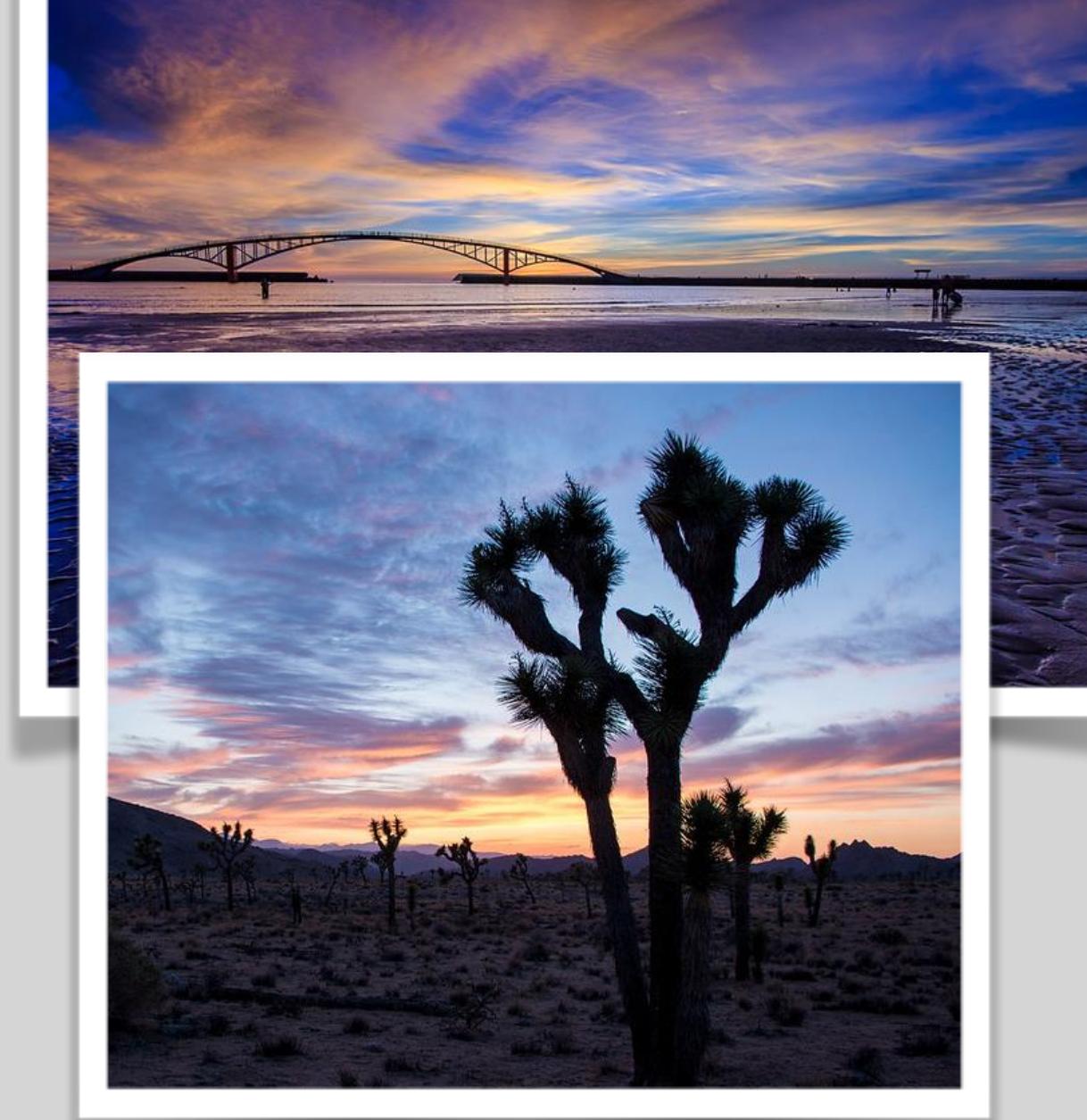
(255,0,0)

white white white white white white white white white green white white white white white
white white white white white white white white green green green white red red white
white white white white red red red red green green brown red red red red
white white white red red red red red brown green red red red red red
white green brown red
green green brown red
white green green brown red white
white white white white red red red black red red red red red white white
white white white white white white white white red red red red white white white
white white white white white white white white white red red white white white white



1 Billion sunsets

slide from Steve Seitz's [video](#)



white

Large Language Model

A
image
raspberry

slide from Steve Seitz's [video](#)

white

Large Language Model

A image white
raspberry

slide from Steve Seitz's [video](#)

red

large Language Model

A image white white
raspberry

slide from Steve Seitz's [video](#)

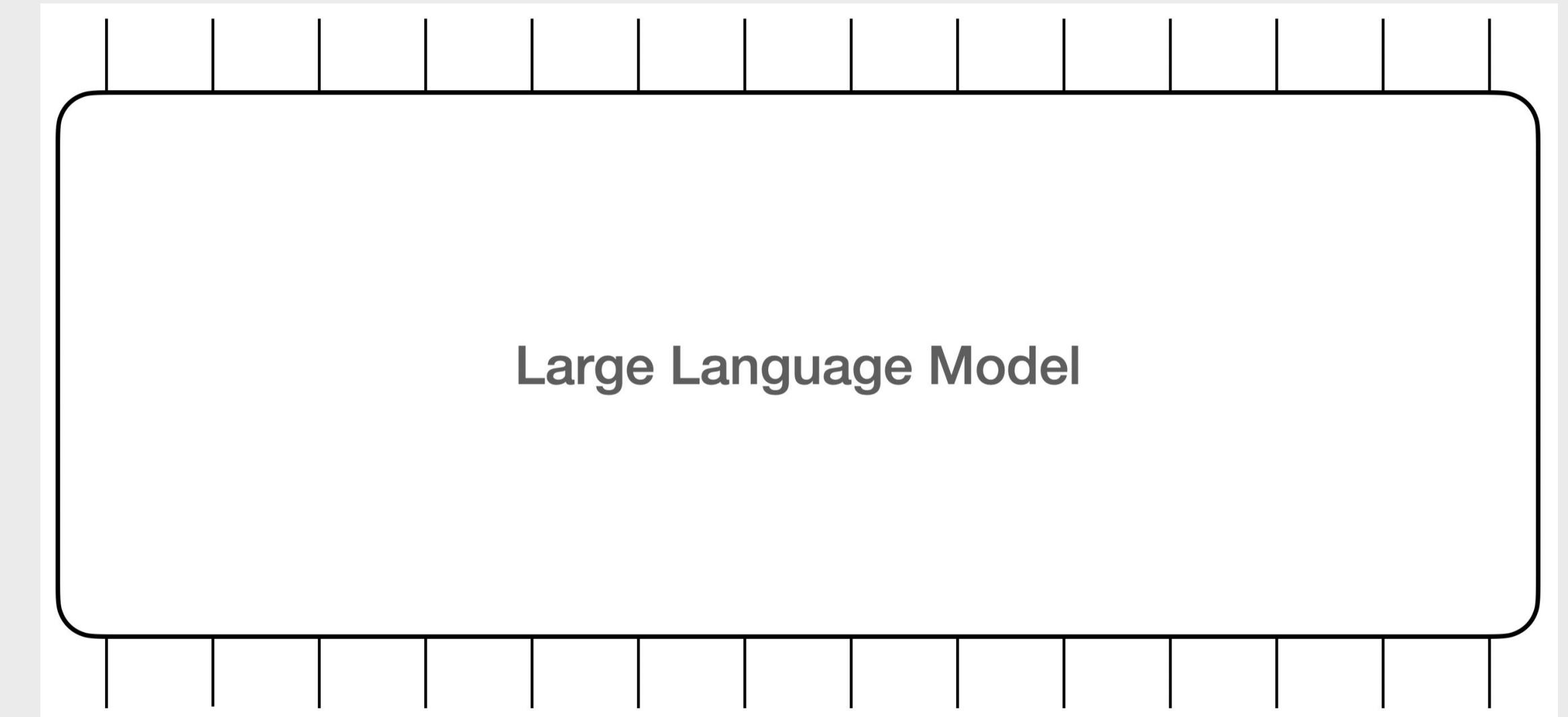
Large Language Model

A image white white red red red white white green green green white
raspberry

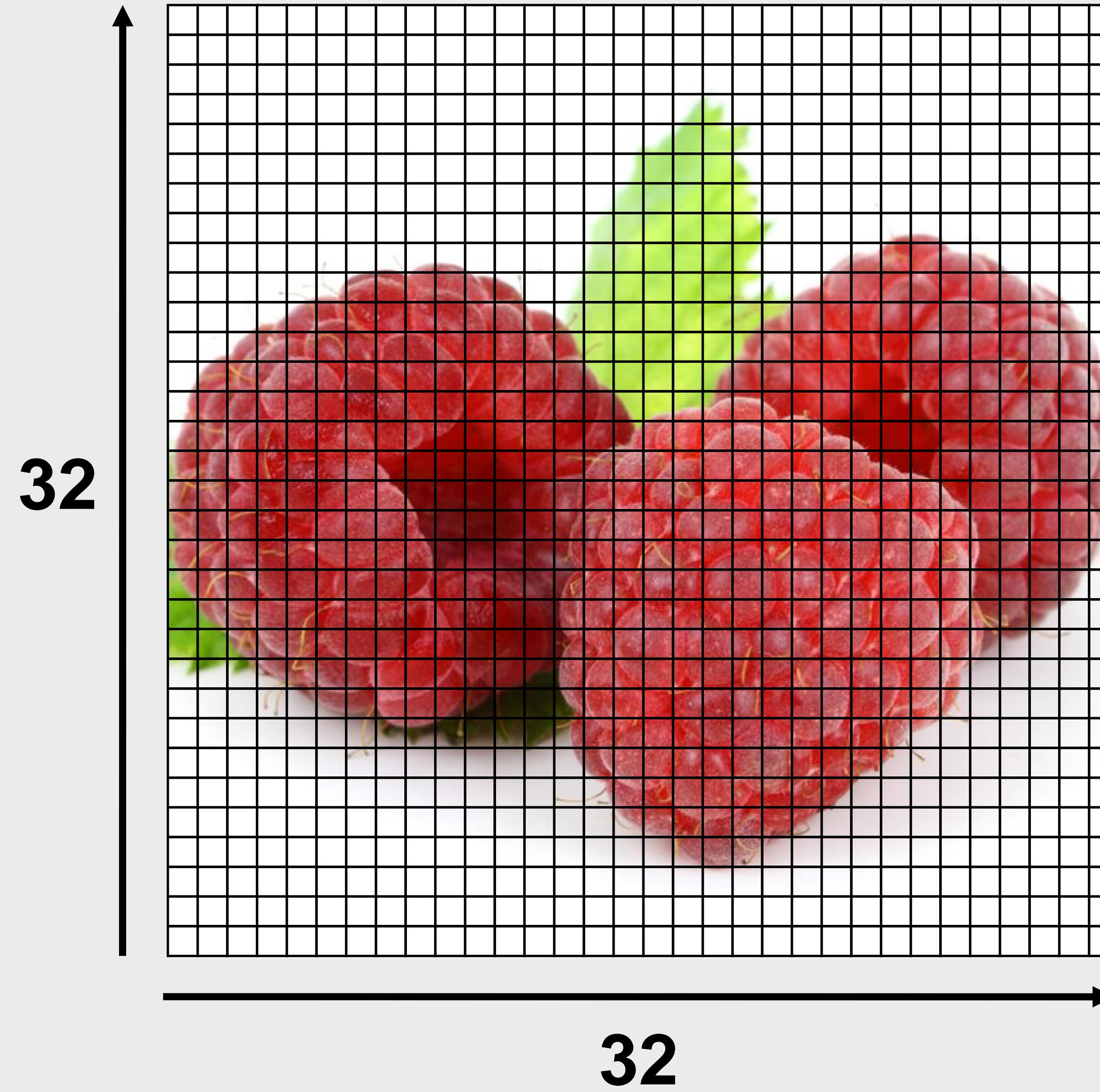
slide from Steve Seitz's [video](#)



1,000,000s of pixels



1,000s of words



slide from Steve Seitz's [video](#)

Torralba et al, 80 Million Tiny Images (2006)

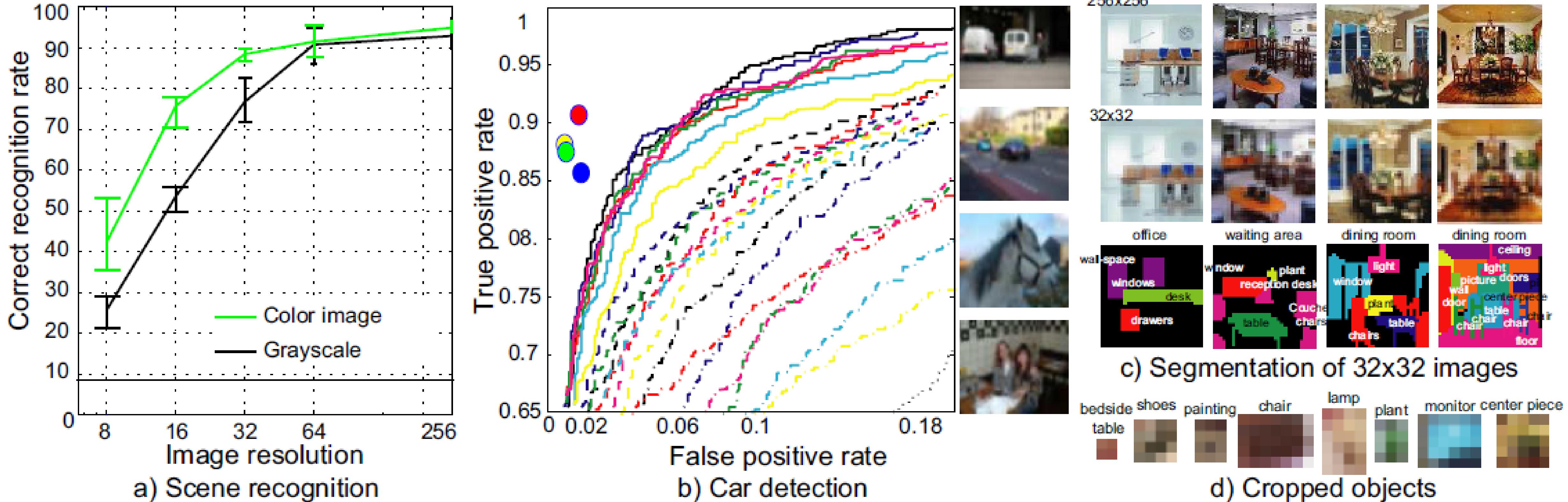
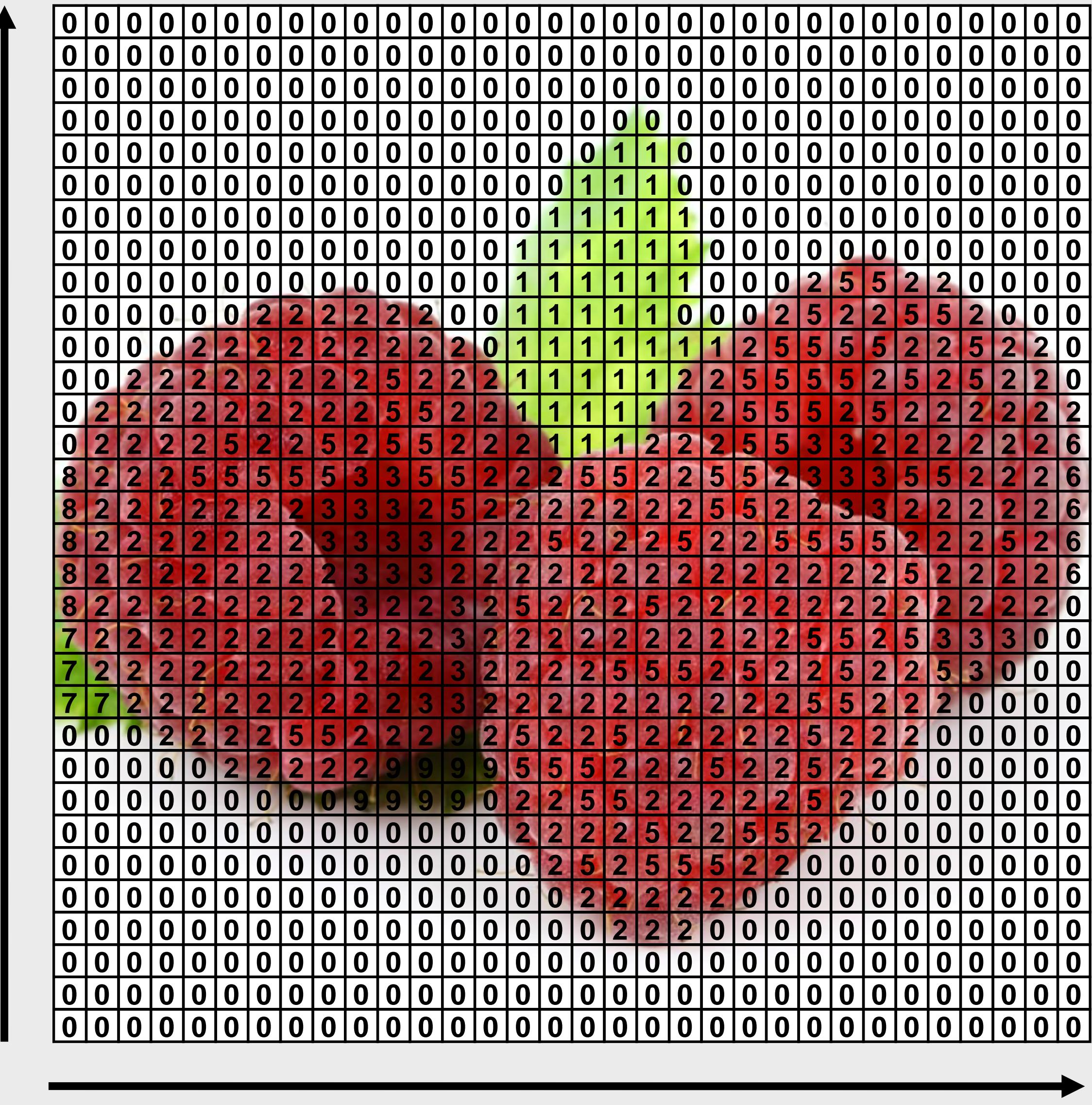


Fig. 1. a) Human performance on scene recognition as a function of resolution. The green and black curves show the performance on color and gray-scale images respectively. For color 32×32 images the performance only drops by 7% relative to full resolution, despite having 1/64th of the pixels. b) Car detection task on the PASCAL 2006 test dataset. The colored dots show the performance of four human subjects classifying tiny versions of the test data. The ROC curves of the best vision algorithms (running on full resolution images) are shown for comparison. All lie below the performance of humans on the tiny images, which rely on none of the high-resolution cues exploited by the computer vision algorithms. c) Humans can correctly recognize and segment objects at very low resolutions, even when the objects in isolation can not be recognized (d).

32

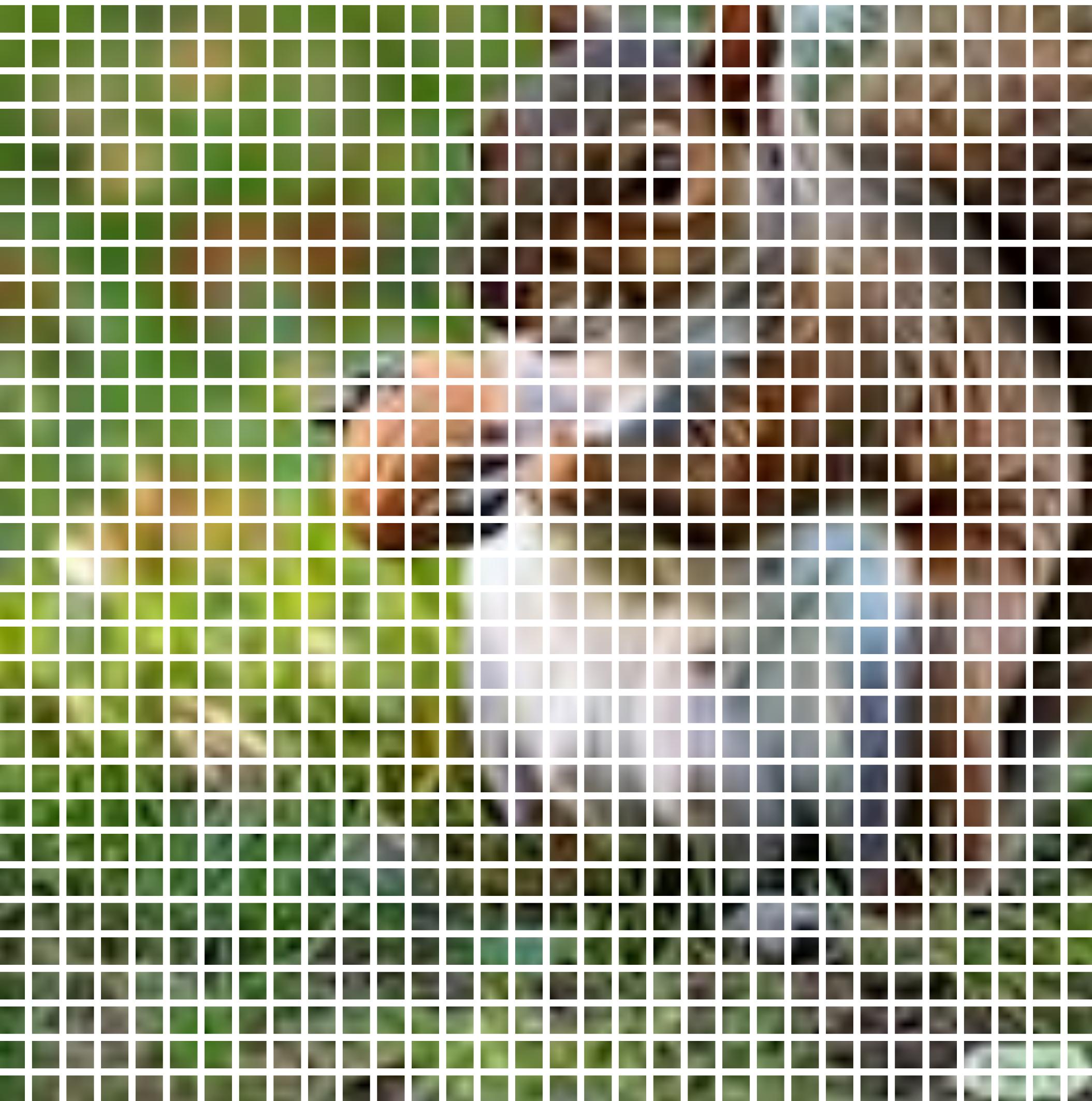


32

$$32 \times 32 = 1024$$

Visual words

squirrel reaching for a nut



squirrel reaching for a nut

slide from Steve Seitz's [video](#)



squirrel reaching for a nut

slide from Steve Seitz's [video](#)

**Up-sampled
4x**



squirrel reaching for a nut

slide from Steve Seitz's [video](#)



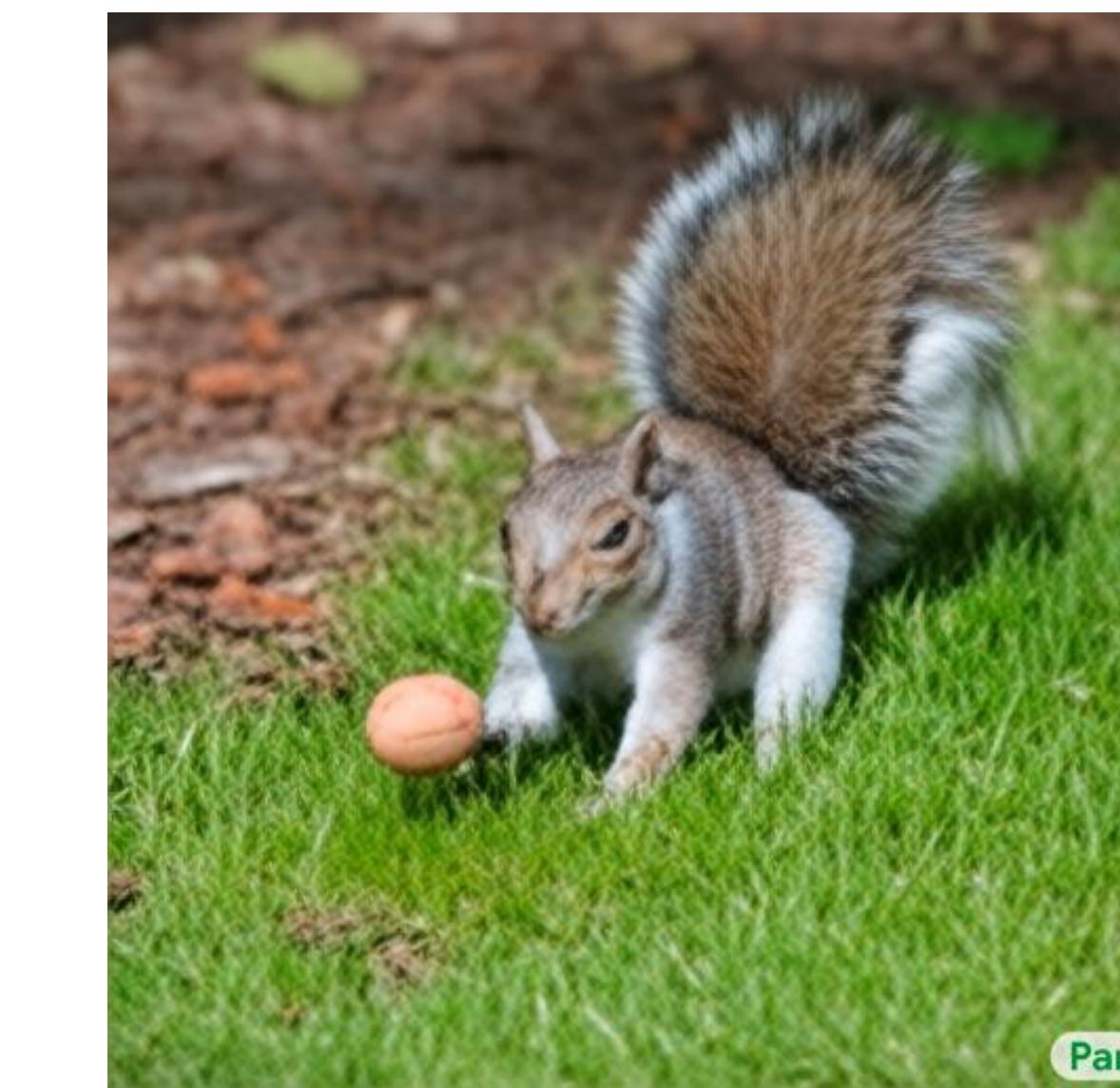
Parti



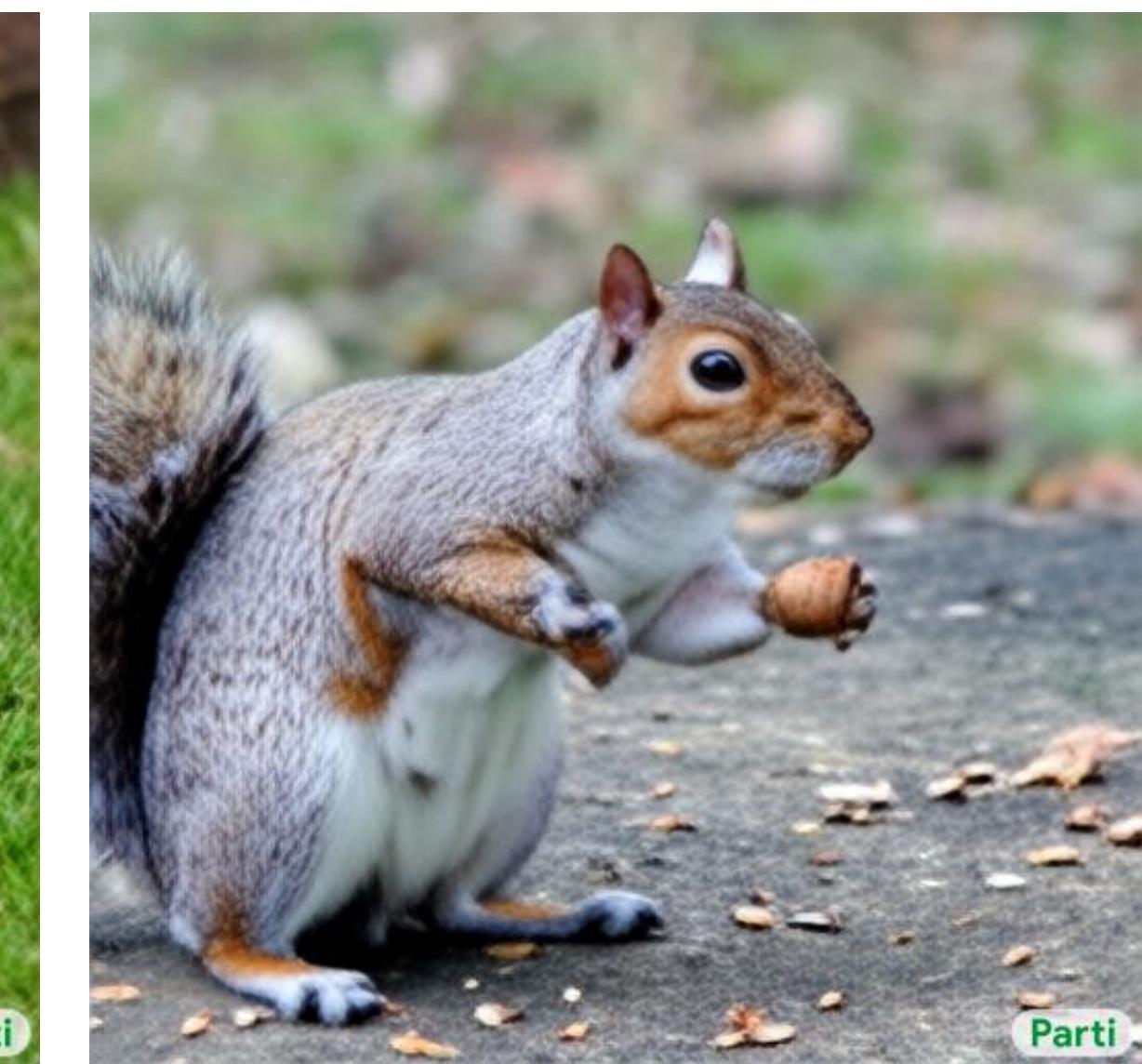
Parti



Parti



Parti



Parti

squirrel reaching for a nut

Parti, <https://parti.research.google/>



Parti



Parti

squirrel reaching for a nut underwater

slide from Steve Seitz's [video](#)



fossil of a squirrel reaching for a nut



slide from Steve Seitz's [video](#)



Parti



Parti

squirrel made of toothpicks wearing sunglasses reaching for a nut

slide from Steve Seitz's [video](#)



Parti

DLSR photograph of a whimsical fantasy house shaped like a squirrel
with windows and a door, in the forest



Parti

slide from Steve Seitz's [video](#)



Squirrel reaching for a nut. by Leonardo da Vinci

slide from Steve Seitz's [video](#)



Squirrel reaching for a nut. Van Gogh painting

slide from Steve Seitz's [video](#)



Part I



Part II

Intricately carved cathedral door of a squirrel reaching for a nut

slide from Steve Seitz's [video](#)



Squirrel reaching for a nut. Woodcut tessellation pattern by M.C. Escher

slide from Steve Seitz's [video](#)



Squirrel reaching for a nut. Latte art



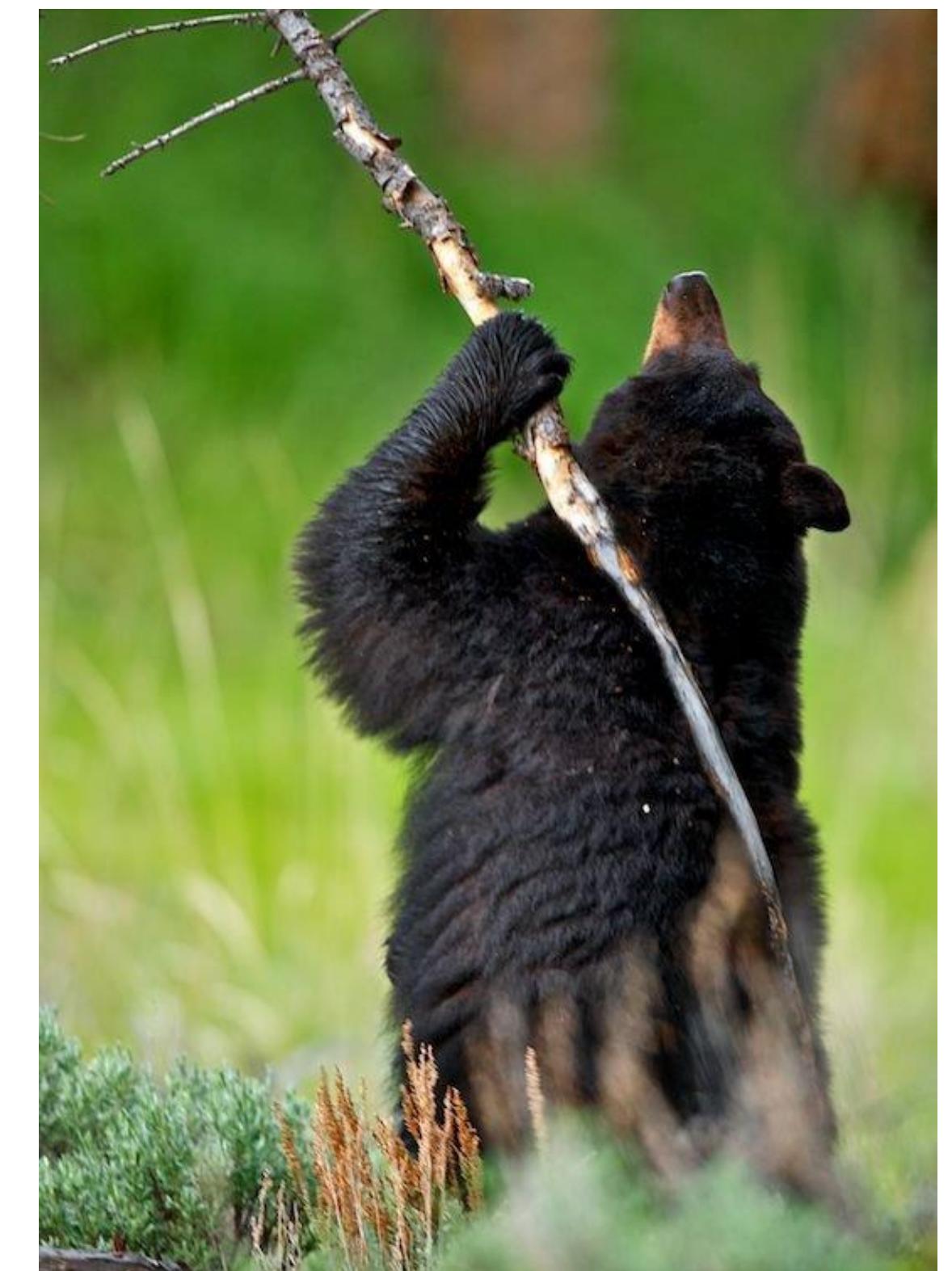
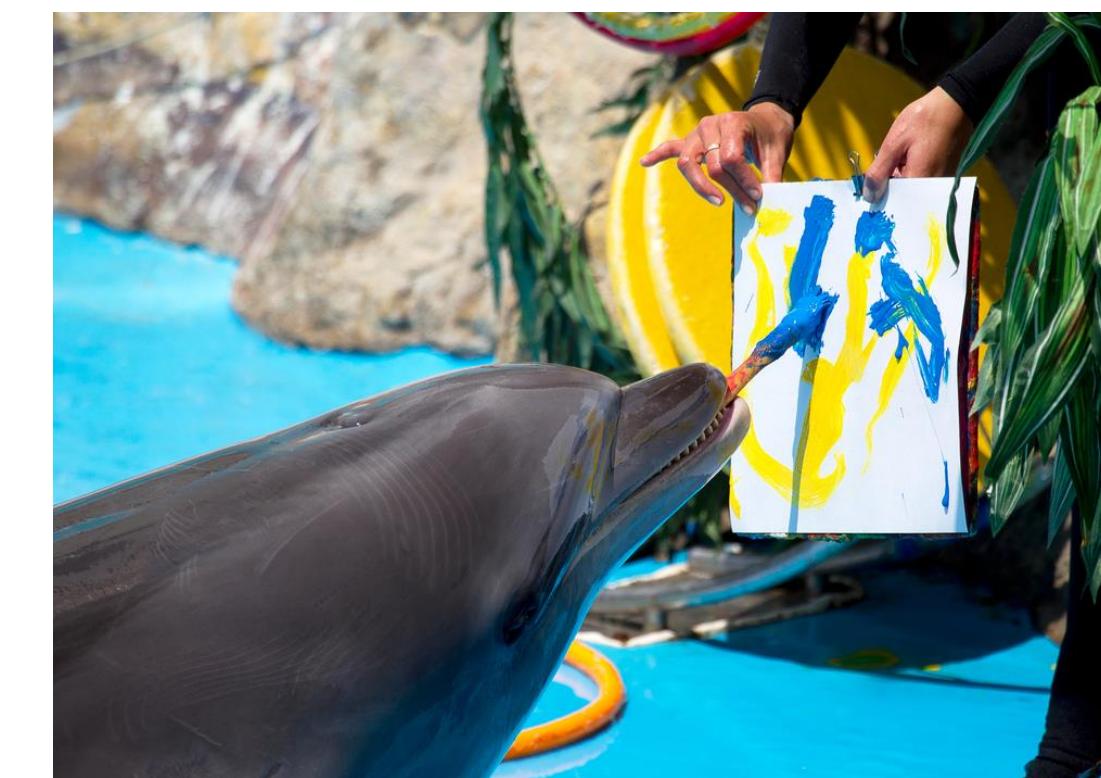
slide from Steve Seitz's [video](#)

Sequential Modeling Enables Scalable Learning for Large Vision Models

Yutong Bai, Xinyang Geng,
Karttikeya Mangalam, Amir Bar,
Alan Yuille, Trevor Darrell, Jitendra Malik, Alexei A Efros

In CVPR'24

Top-down vs. bottom-up



Scientific Question:
How far can we go from pixels **alone**?

across:
images,
videos,
supervised / unsupervised
synthetic / real,
all kinds of tasks
2D / 3D / 4D data etc.

Dataset	Tokens (Millions)	Annotation Type	Annotation Source
Unpaired Image Data			
LAION 5B [71] (1.5B images subset)	380690	-	-
Images with Annotations			
ImageNet 1K [25]	1317.40	Image Classification	Ground Truth
COCO [54]	363	Object Detection	MMDetection [16]
ADE 20K [100], Cityscapes [22]	66.88	Semantic Segmentation	Ground Truth
COCO [54], ImageNet 1K [25]	2078.06	Semantic Segmentation	Mask2Former [19]
COCO [54], lvmhp [51], mpii [4], Unite [49]	950.79	Human Pose	MMPose[21]
COCO [54], ImageNet 1K [25]	1623.85	Depth Map Image	DPT [67]
Subset of InstructPix2Pix [34]	415.46	Style Transfer	InstructPix2Pix [34]
COCO[54], ImageNet 1K[25]	1623.85	Surface Normal Image	NLL-AngMF [7]
COCO [54], ImageNet 1K [25]	1623.85	Edge Detection	DexiNed [79]
DID-MDN [98]	35.06	Rainy and Clean Image Pairs	Ground Truth
SIDD [3]	245.76	Denoised Image	Ground Truth
LOL[89]	0.458	Light Enhanced Image	Ground Truth
ImageNet 1K [25]	1321.07	Grayscale and Colorized Image Pairs	Ground Truth
ImageNet 1K [25]	1321.07	Inpainting	Ground Truth
Kitti [34]	9.21	Stereo	Ground Truth
Videos			
UCF101 [78]	109.11	-	-
DAVIS [65]	0.36	-	-
HMDB [48]	55.41	-	-
ActivityNet [13]	380.63	-	-
Moments in Time [59]	2979.00	-	-
Multi-moments in Time [60]	4124.04	-	-
Co3D [69]	228.75	-	-
Charades v1 [76]	241.53	-	-
Something-something v2 [37]	904.57	-	-
YouCook [23]	3.14	-	-
Kinetics 700 [14]	7092.04	-	-
MSR-VTT [92]	57.34	-	-
Youtube VOS [93]	63.70	-	-
jester [57]	606.47	-	-
diving48 [52]	150.73	-	-
MultiSports [53]	78.44	-	-
CharadesEgo [77]	193.06	-	-
AVA [61]	117.96	-	-
Ego4D [38]	1152.12	-	-
Videos with Annotations			
VIPSeg [58]	64.47	Video Panoptic Segmentation	Ground Truth
Hand14K [32]	1.96	Hand Segmentation	Ground Truth
AVA [61]	122.88	Video Detection	Ground Truth
JHMDB [43]	19.00	Optical Flow	Ground Truth
JHMDB [43]	37.92	Video Human Pose	Ground Truth
Synthetic 3D Views			
Objaverse [24] Rendered Multiviews	217.85	-	-

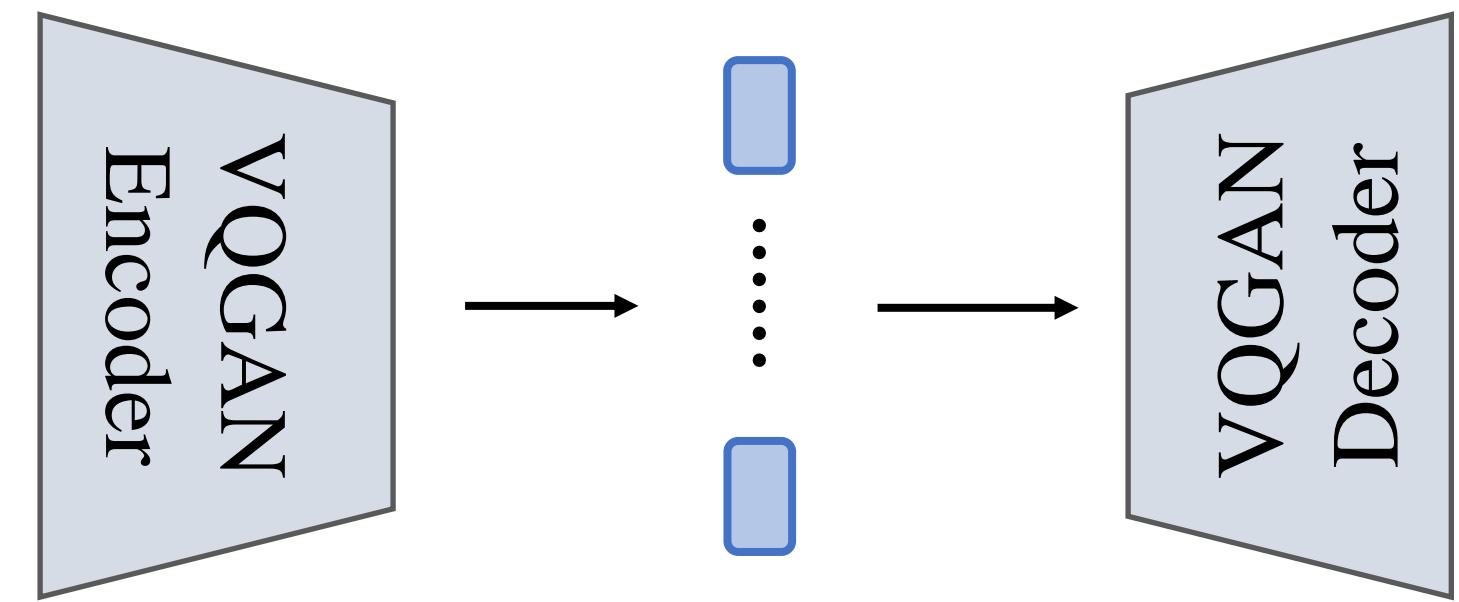
Sentence -> Visual Sentence

Single images



<EOS>

Tokenizer



tok1

tok2

tok3

• • •

Videos

<BOS>



... <EOS>

Image sequences

<BOS>



... <EOS>

categories

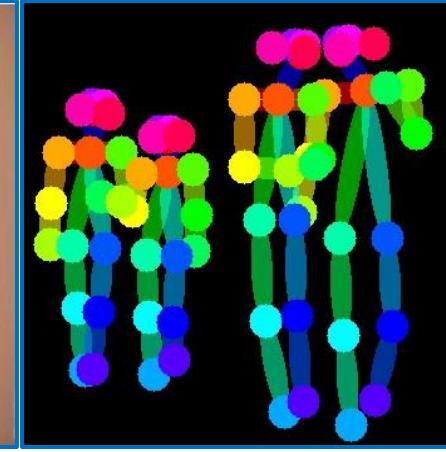
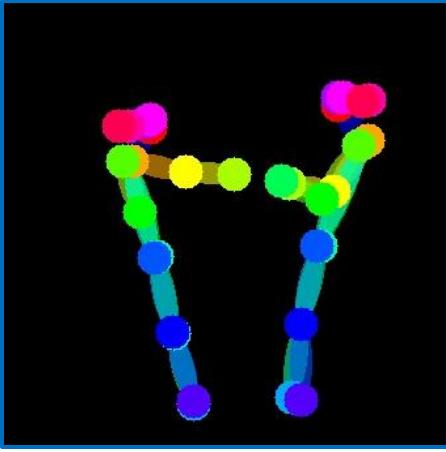
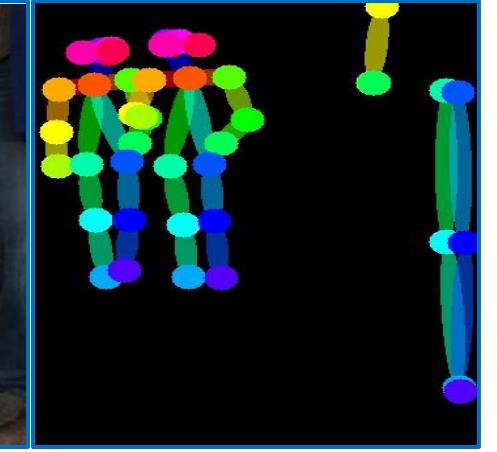
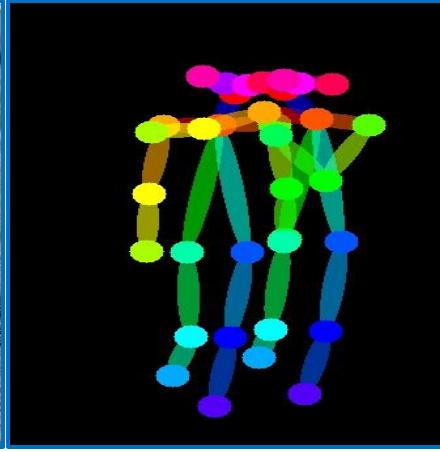
<BOS>



... <EOS>

Images with annotation

<BOS>

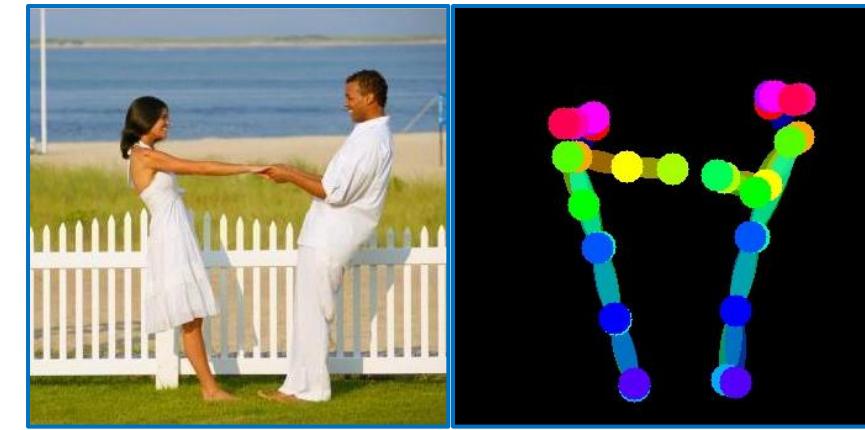


• • • <EOS>

Images with annotation



<BOS>



... <EOS>



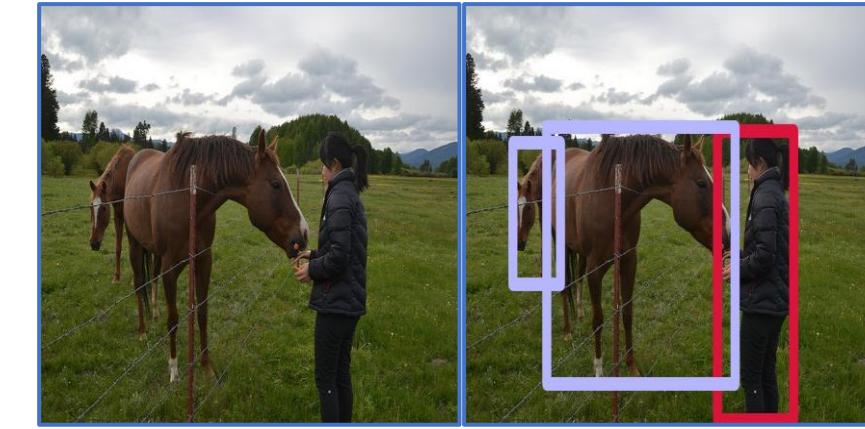
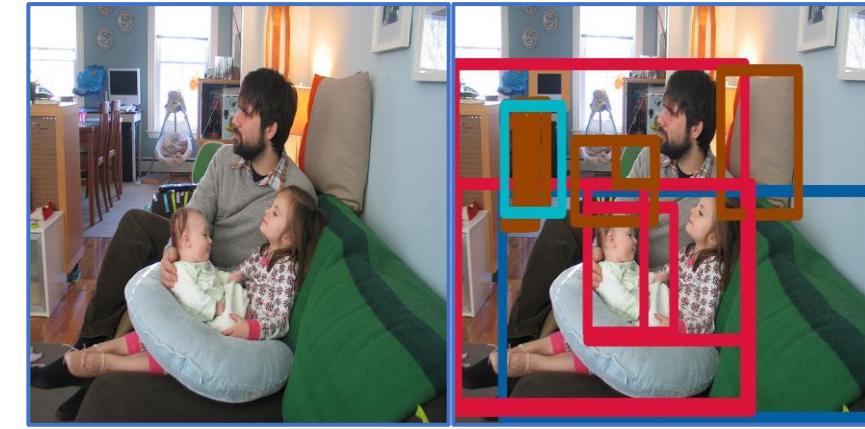
<BOS>



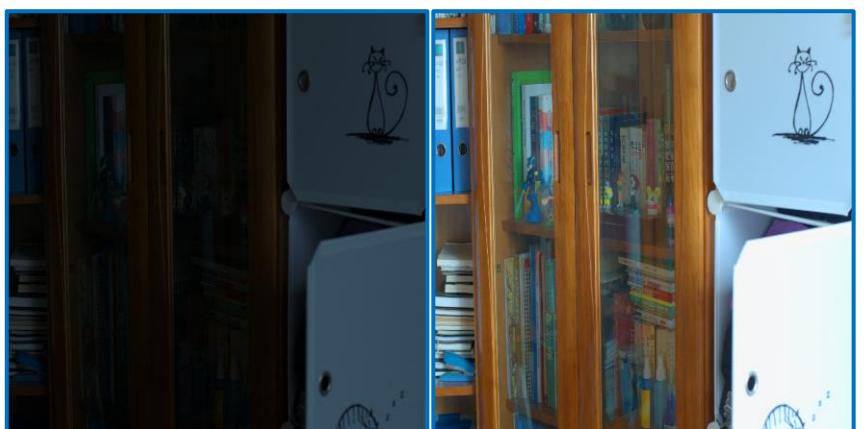
... <EOS>



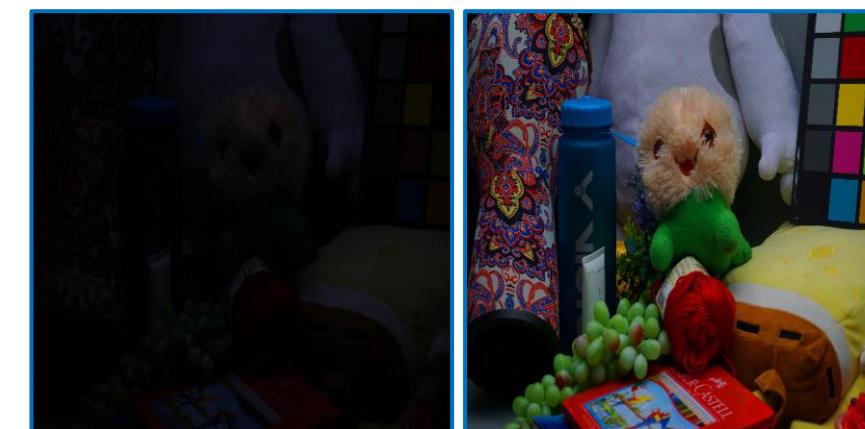
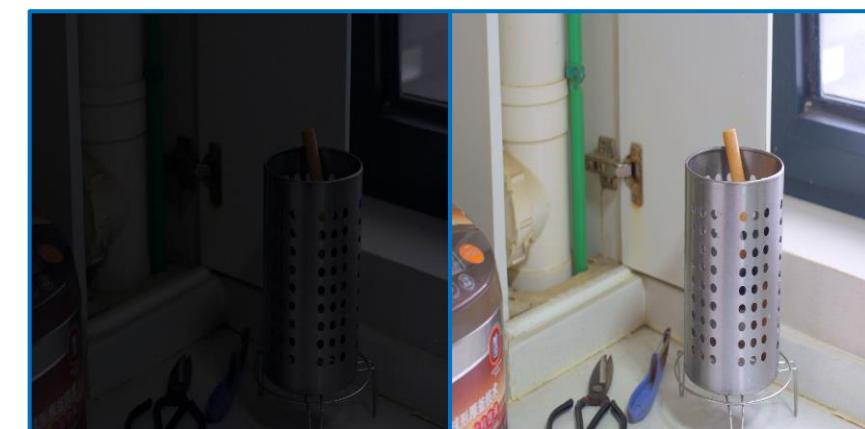
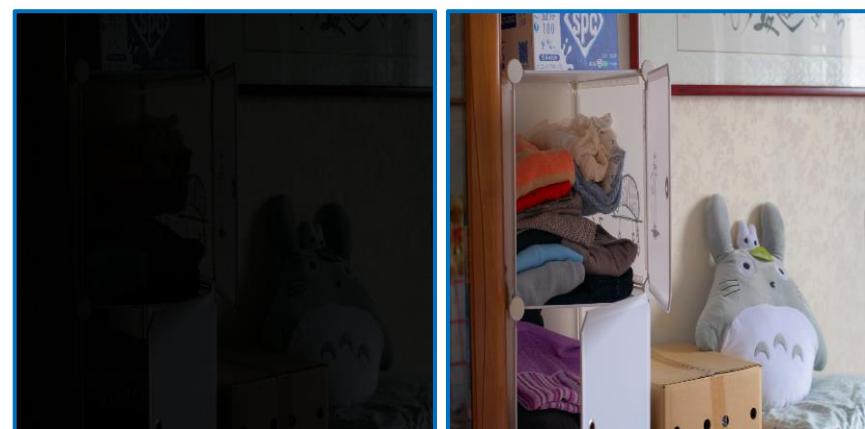
<BOS>



... <EOS>



<BOS>



... <EOS>

Images with free form annotation

<BOS>



...

<EOS>

<BOS>



... <EOS>

<BOS>



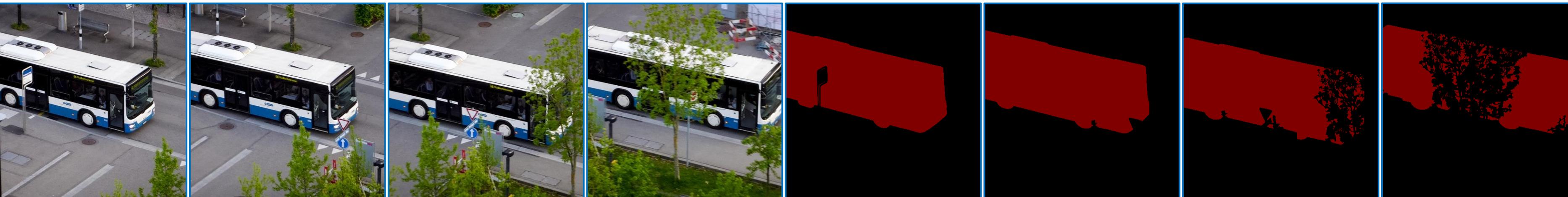
... <EOS>

Videos with annotation

<BOS>

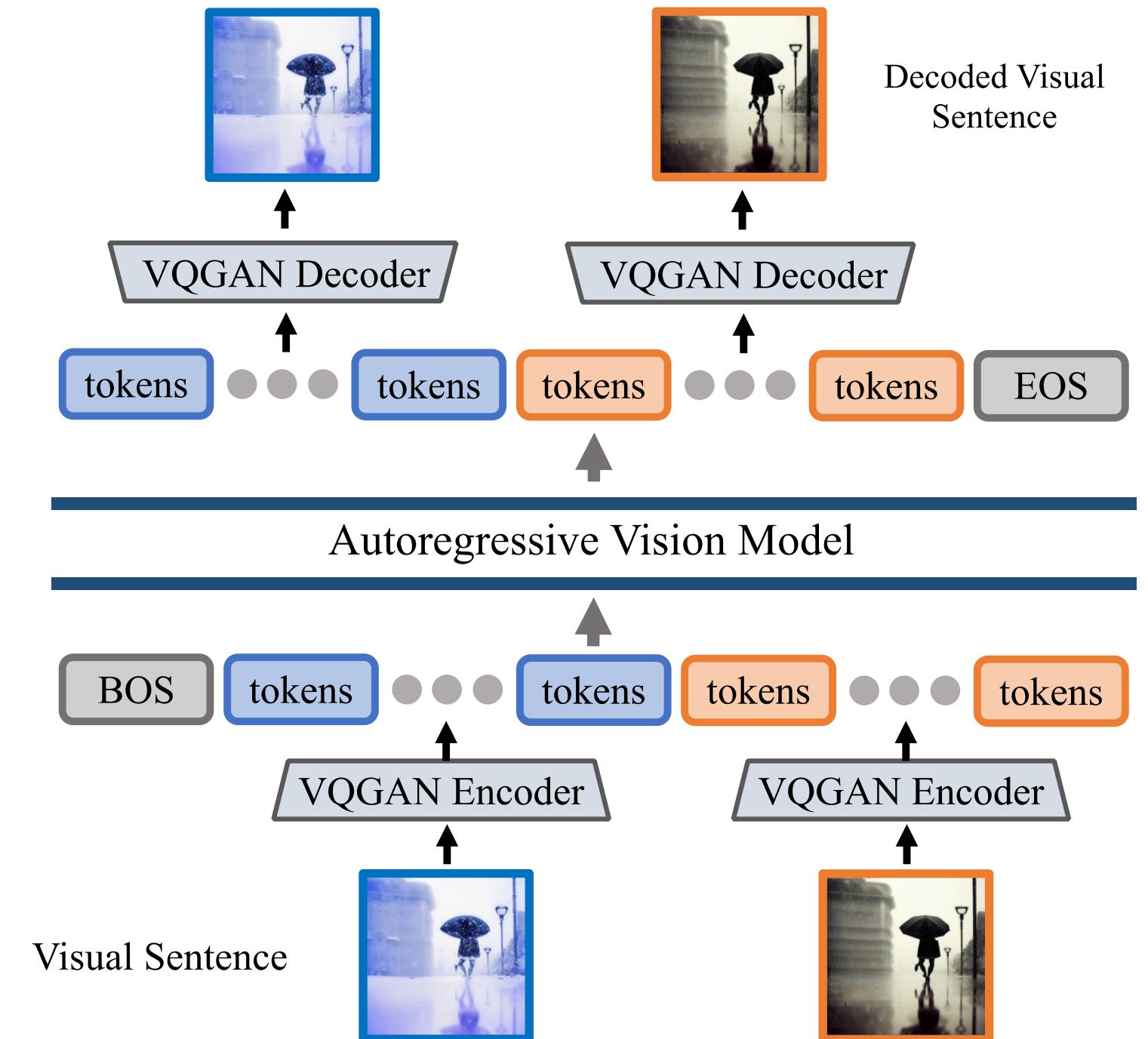
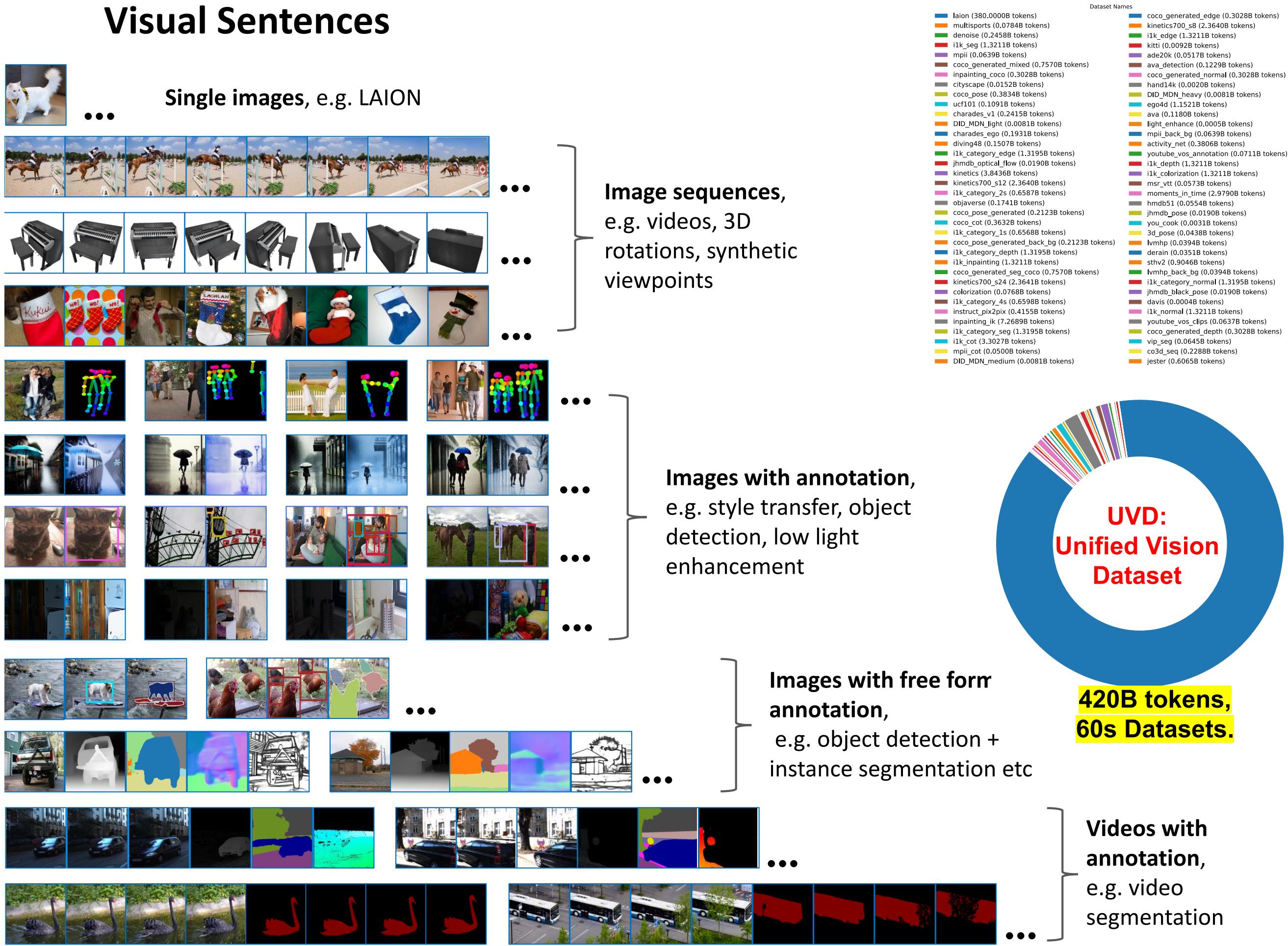


<EOS>

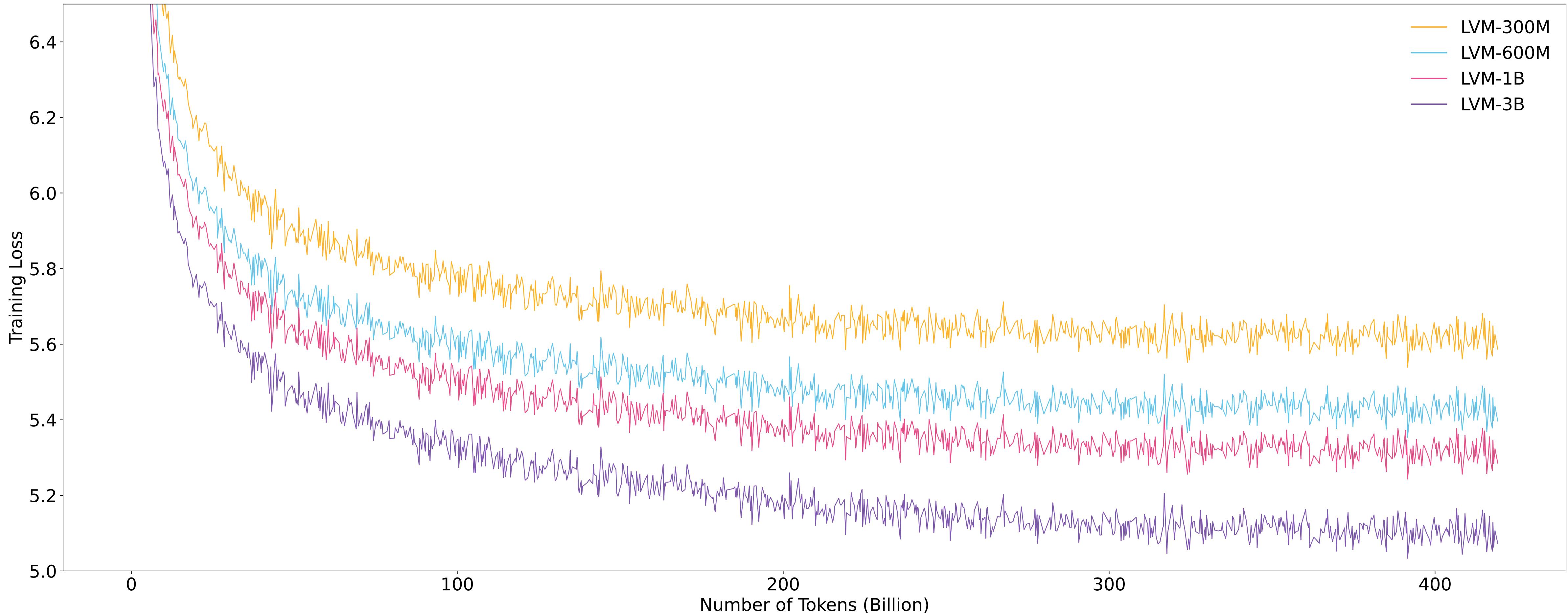


LVM: Large Vision Model

Visual Sentences



Training Loss (1 epoch) ~ Validation Loss



Sequential Prompting

Prompts



Sequential Prompting

Prompts

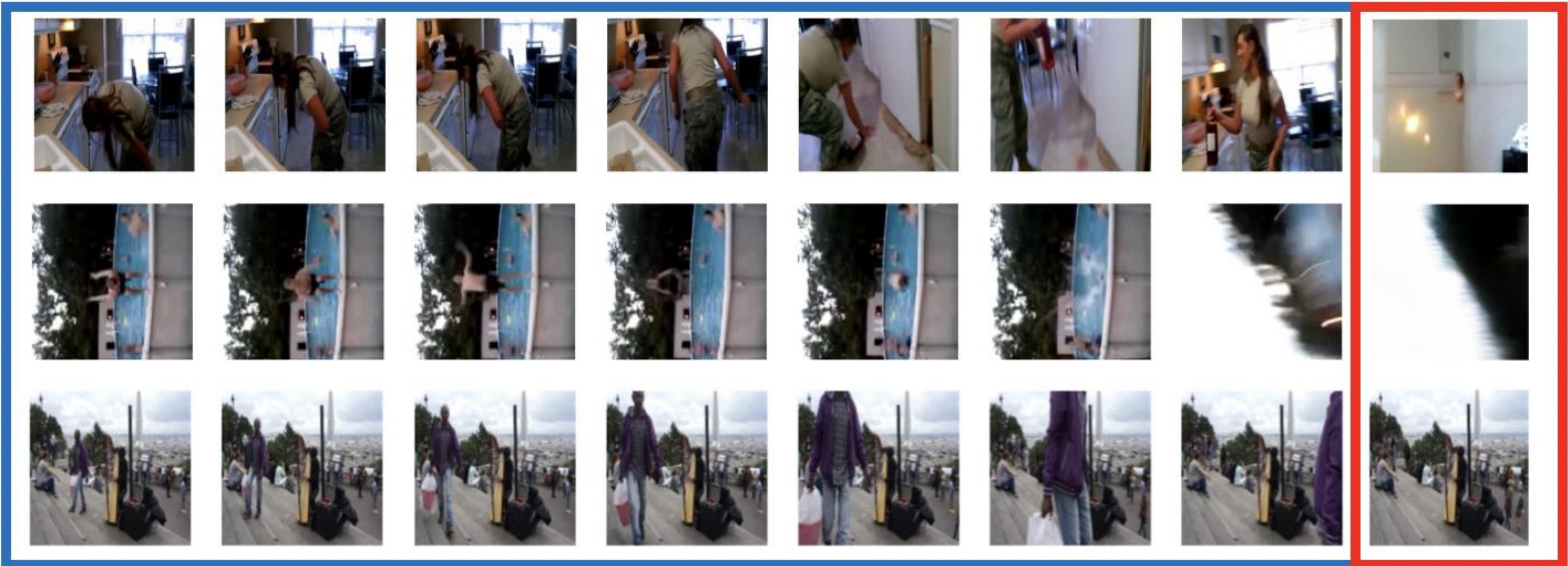
Generated



Sequential Prompting

Prompts

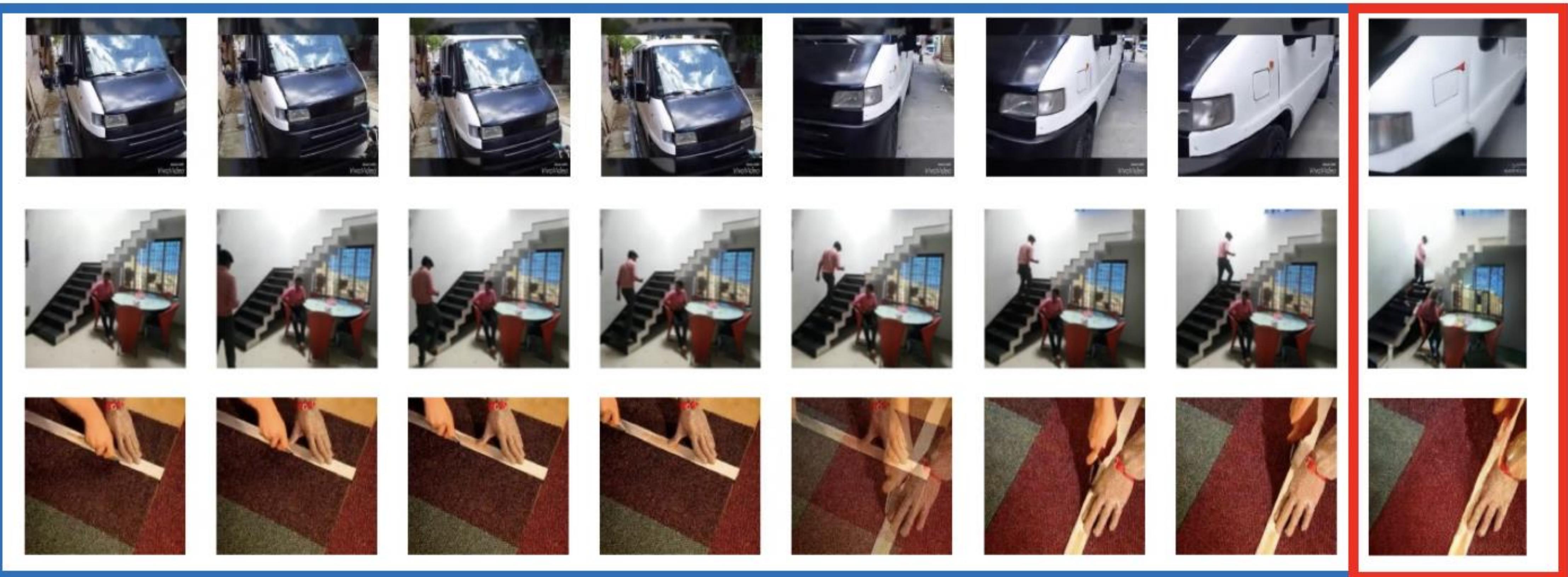
Generated



Sequential Prompting

Prompts

Generated



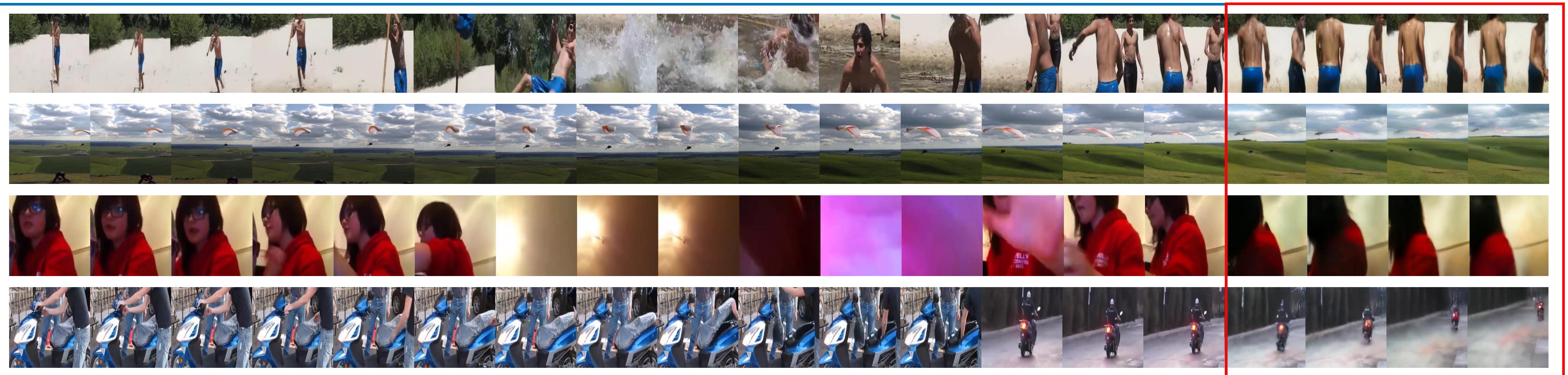
Longer contexts



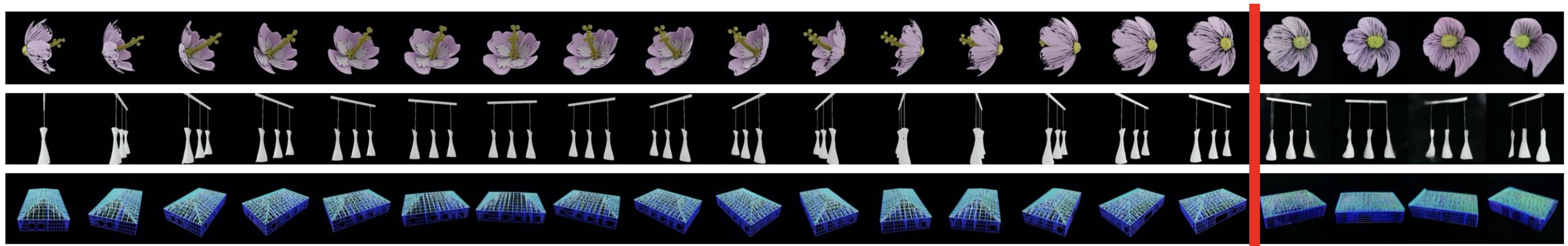
Sequential Prompting

Prompts

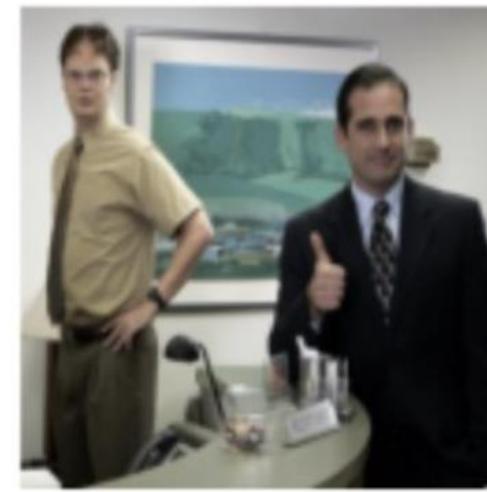
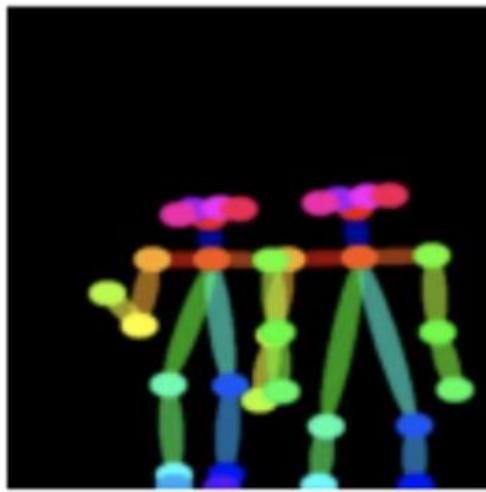
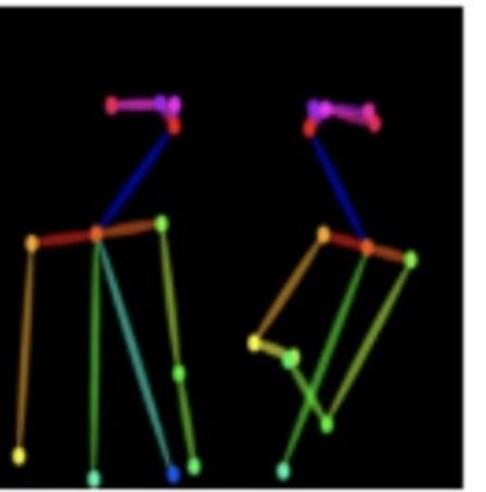
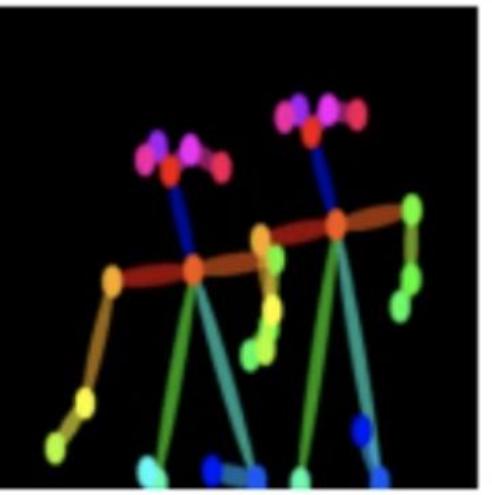
Generated



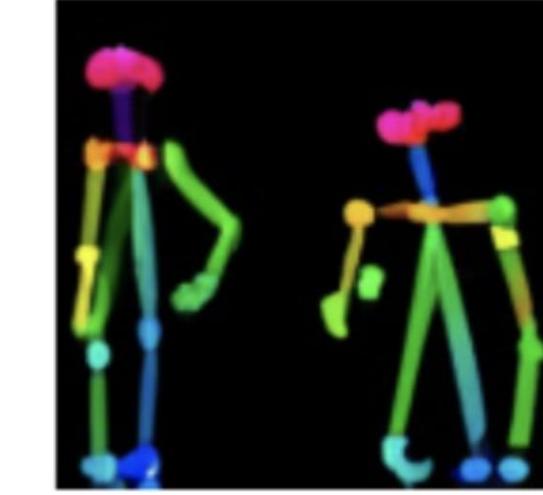
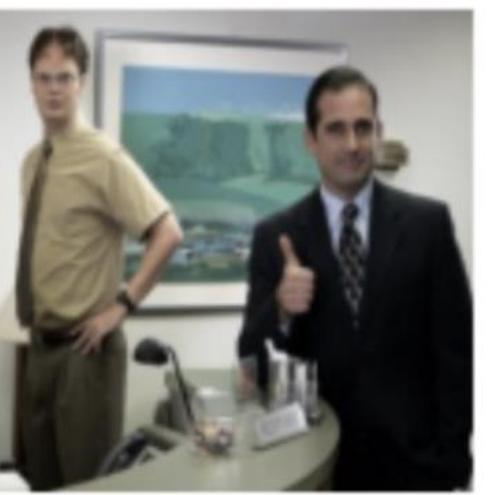
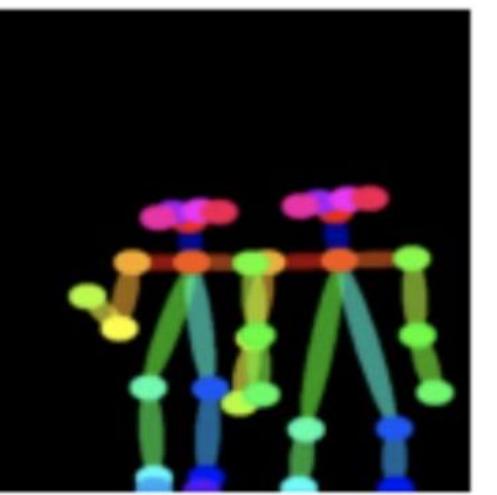
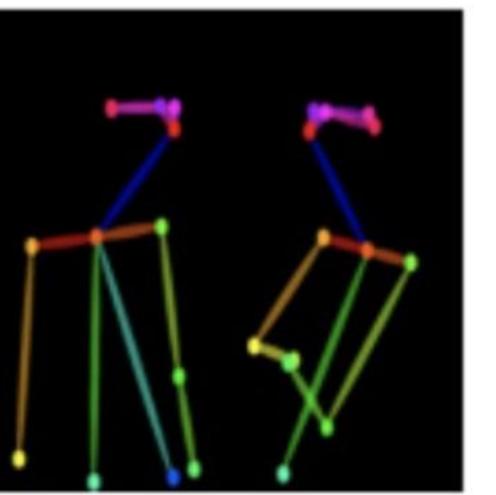
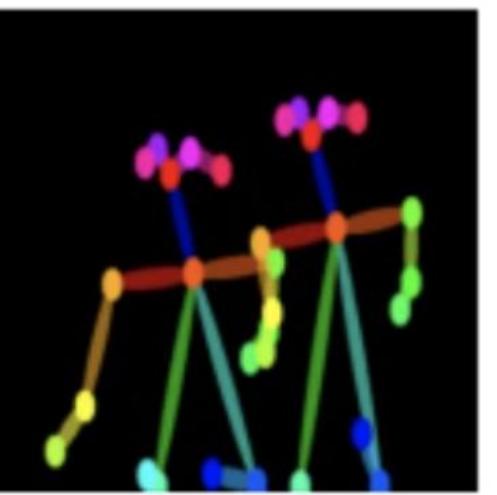
Sequential Prompting



Analogy Prompting

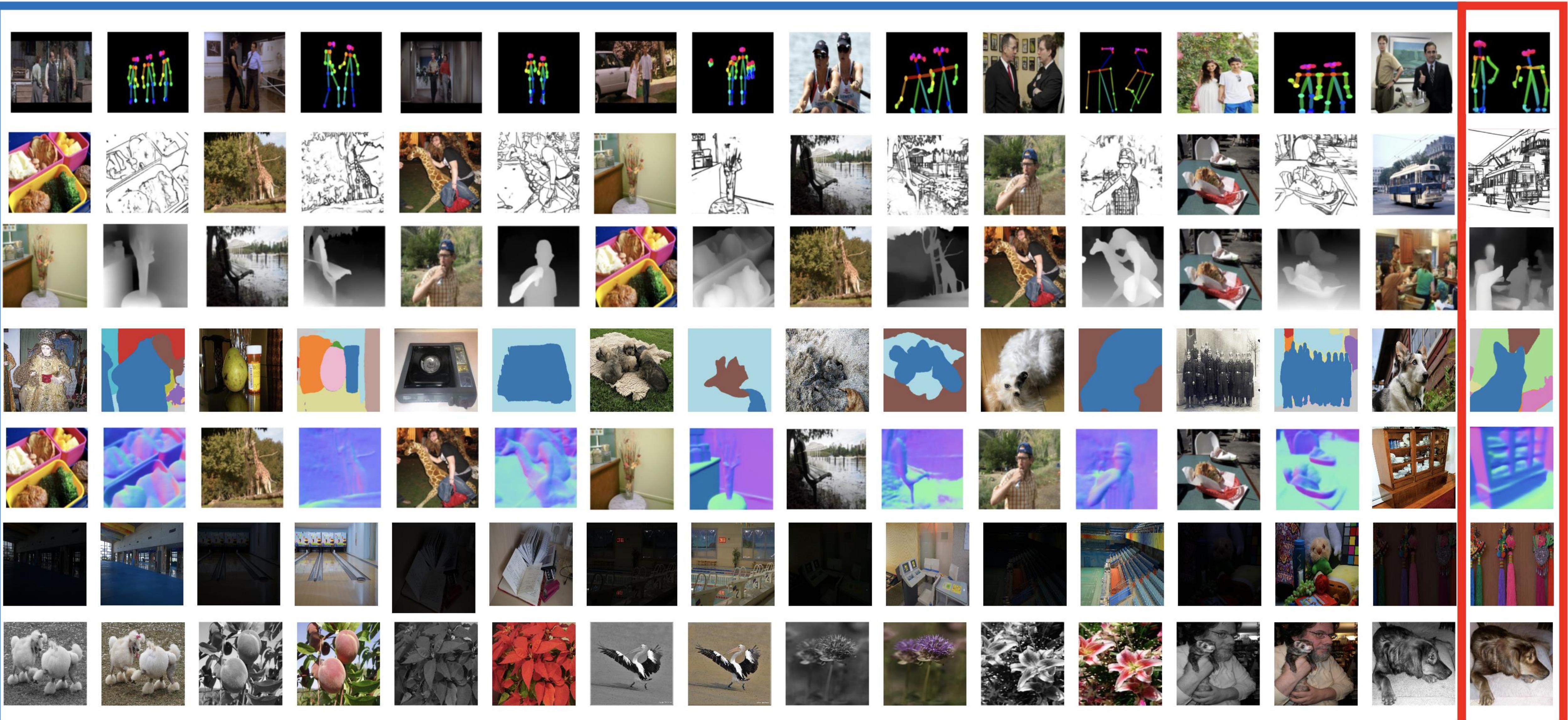


Analogy Prompting

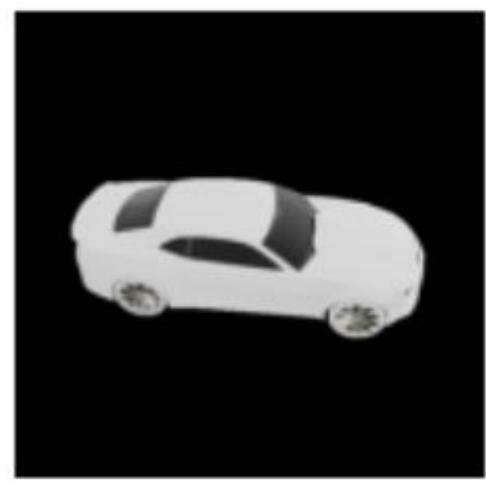


Prompts

Generated



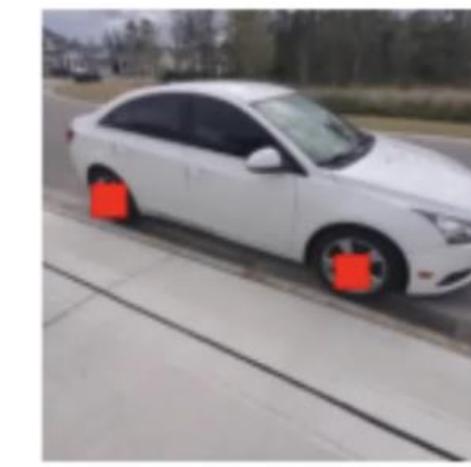
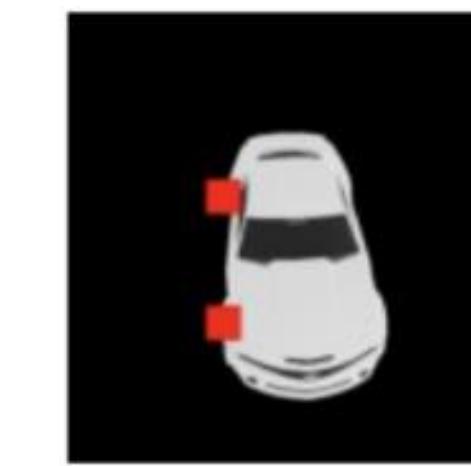
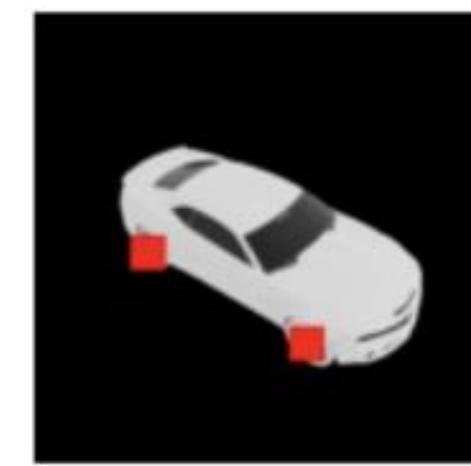
More complicated



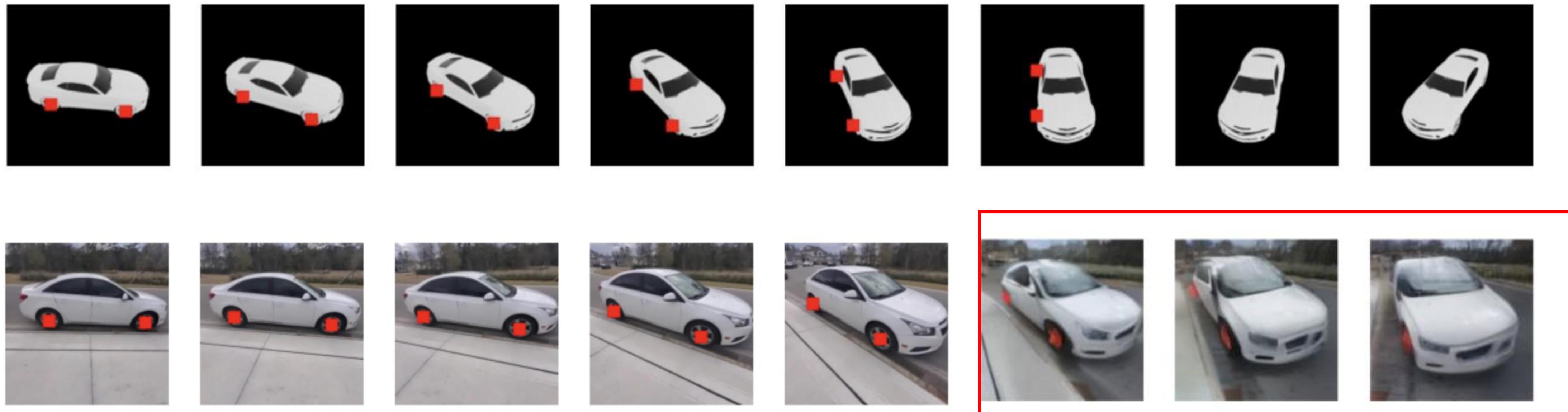
More complicated



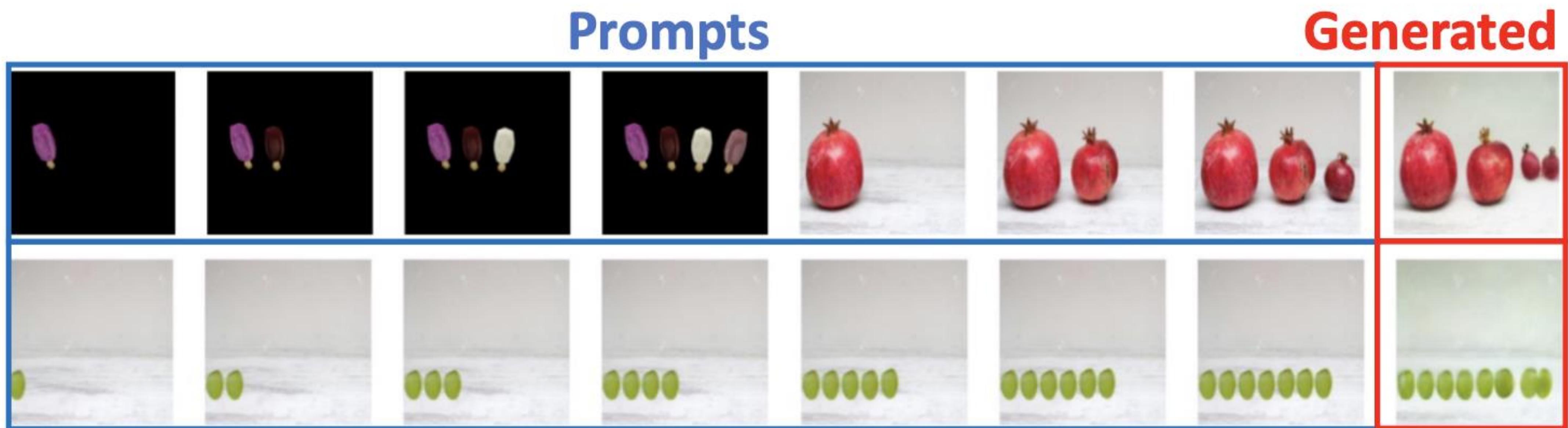
More complicated



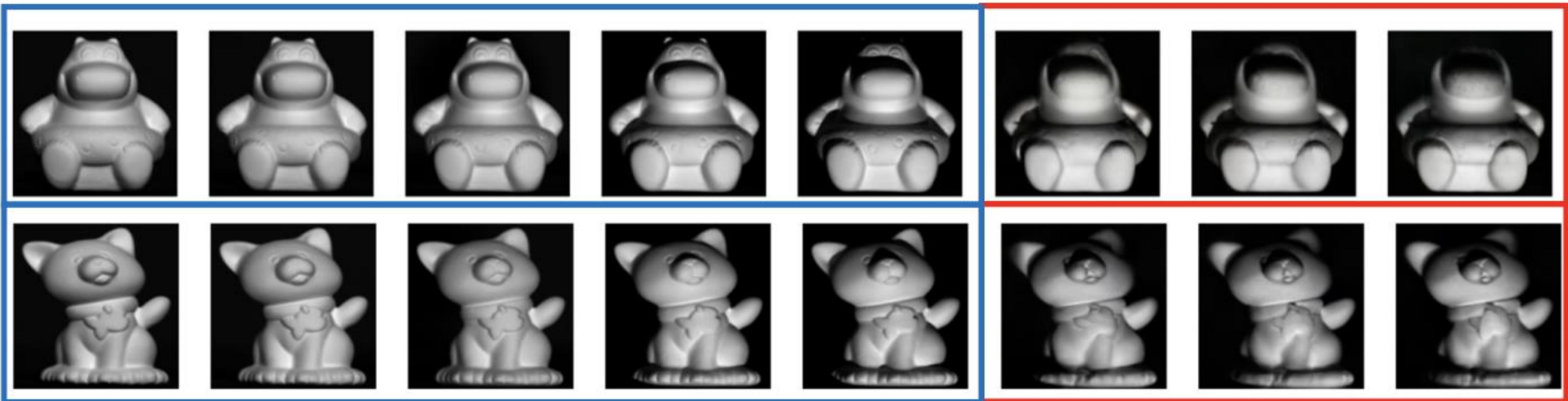
More complicated



Unseen tasks



Unseen tasks



Not easily describable



Not easily describable



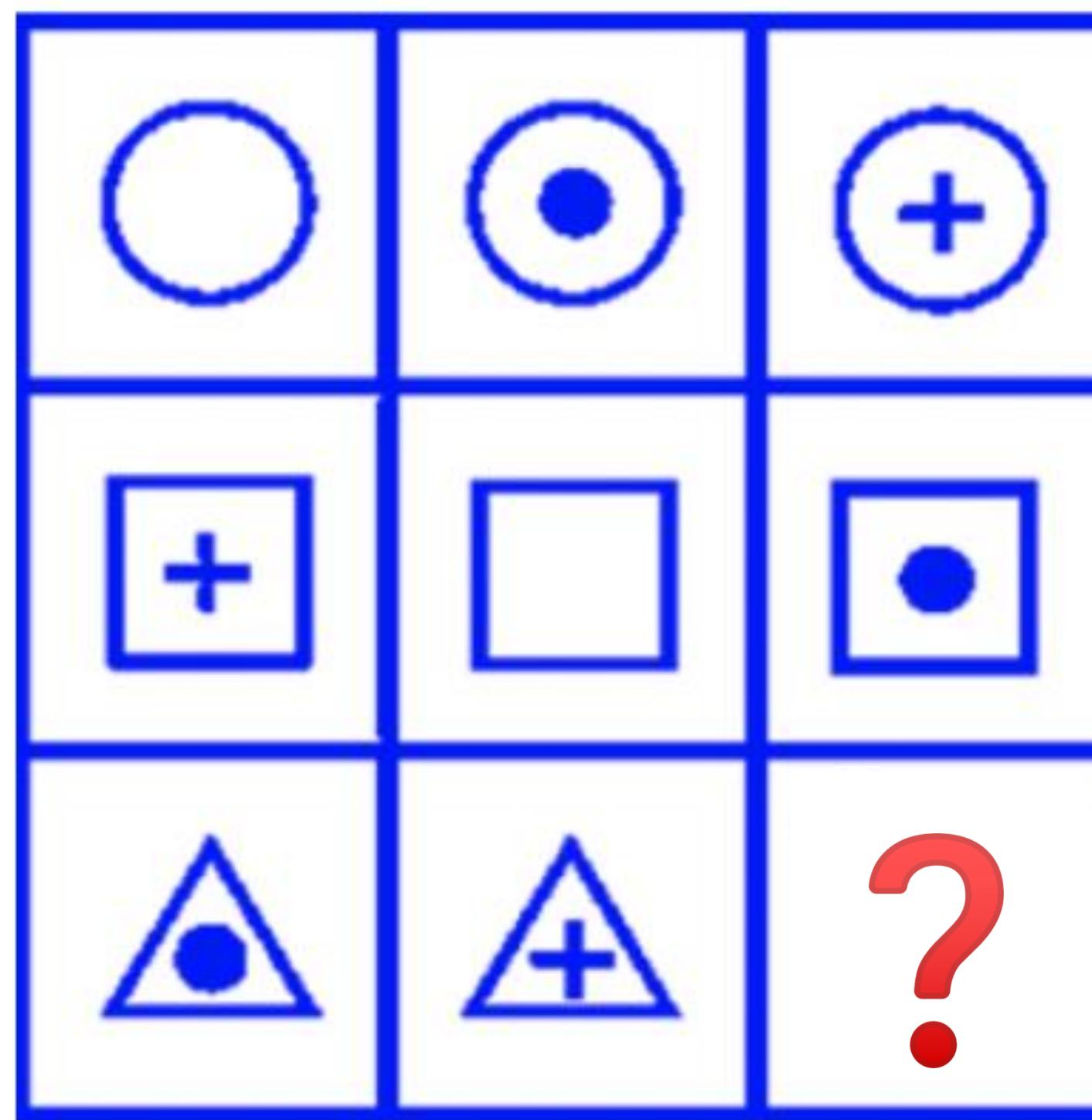
Not easily describable

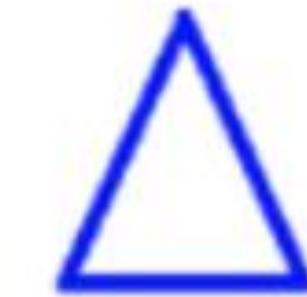
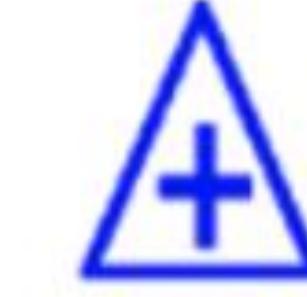
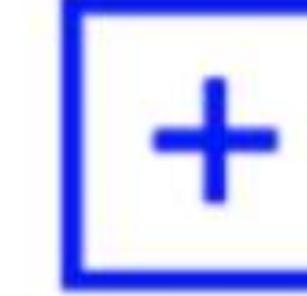
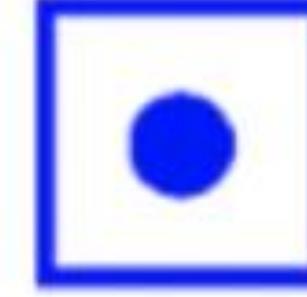
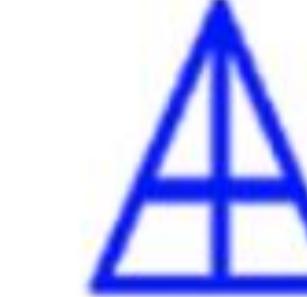


Not easily describable

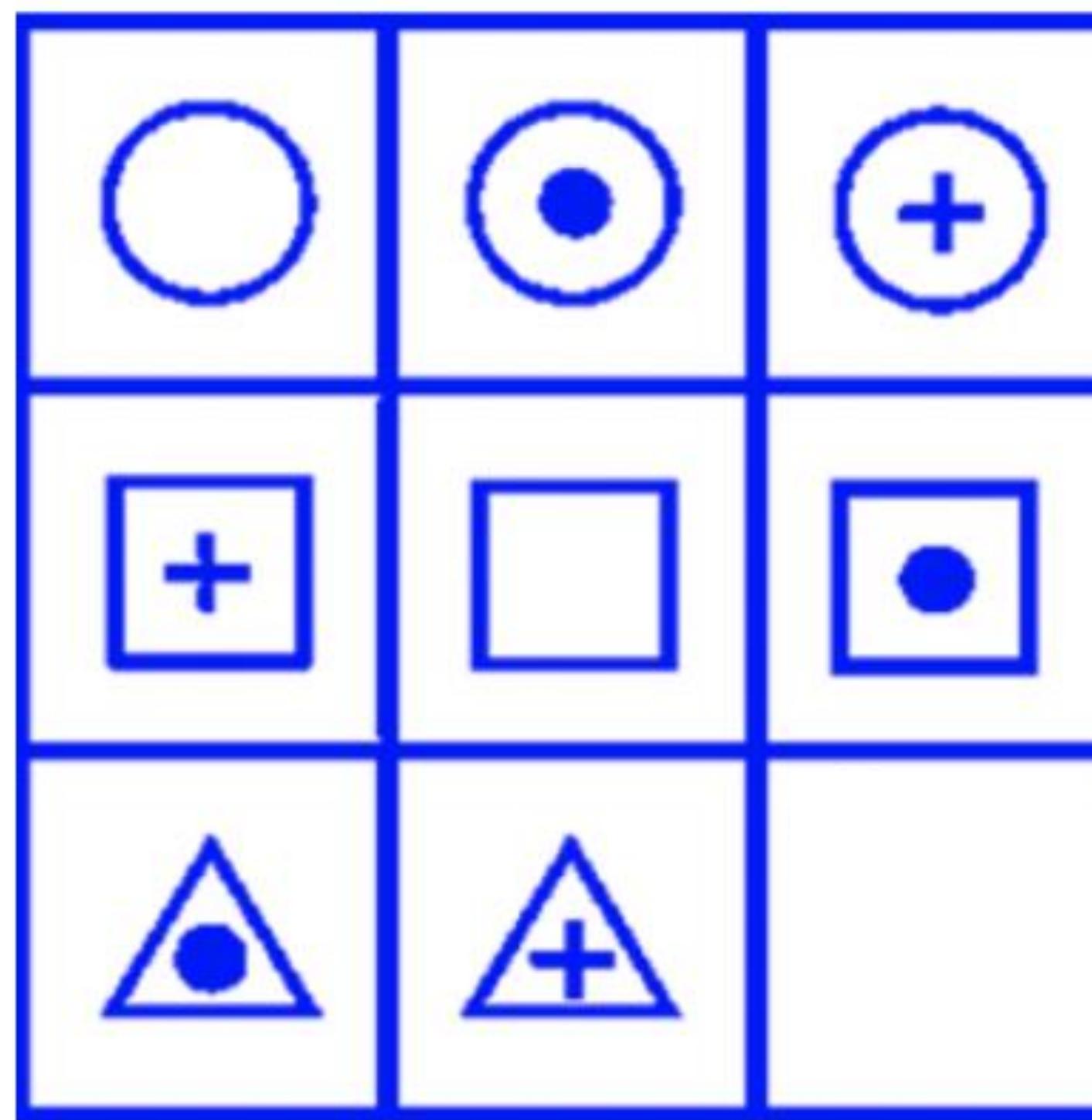


Raven's Progressive Test (Non-verbal IQ test)



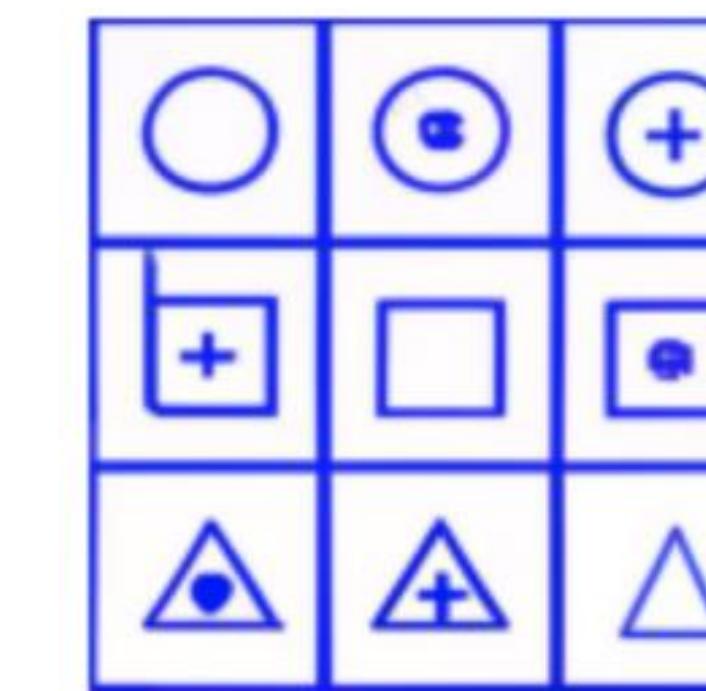
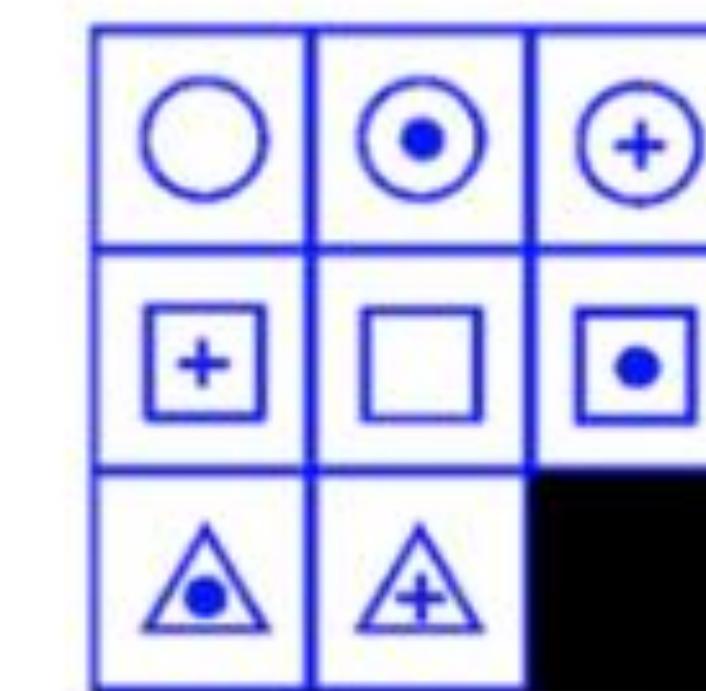
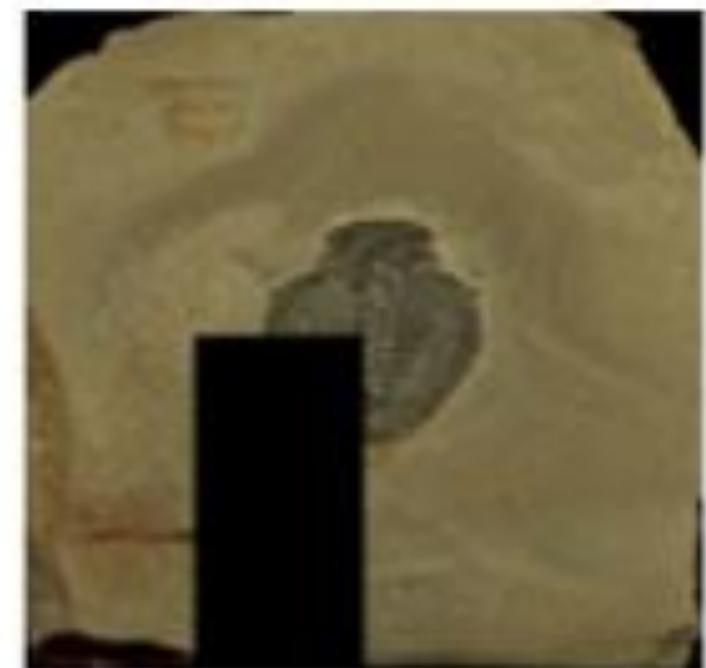
- a) 
- b) 
- c) 
- d) 
- e) 
- f) 

Raven's Progressive Test (Non-verbal IQ test)

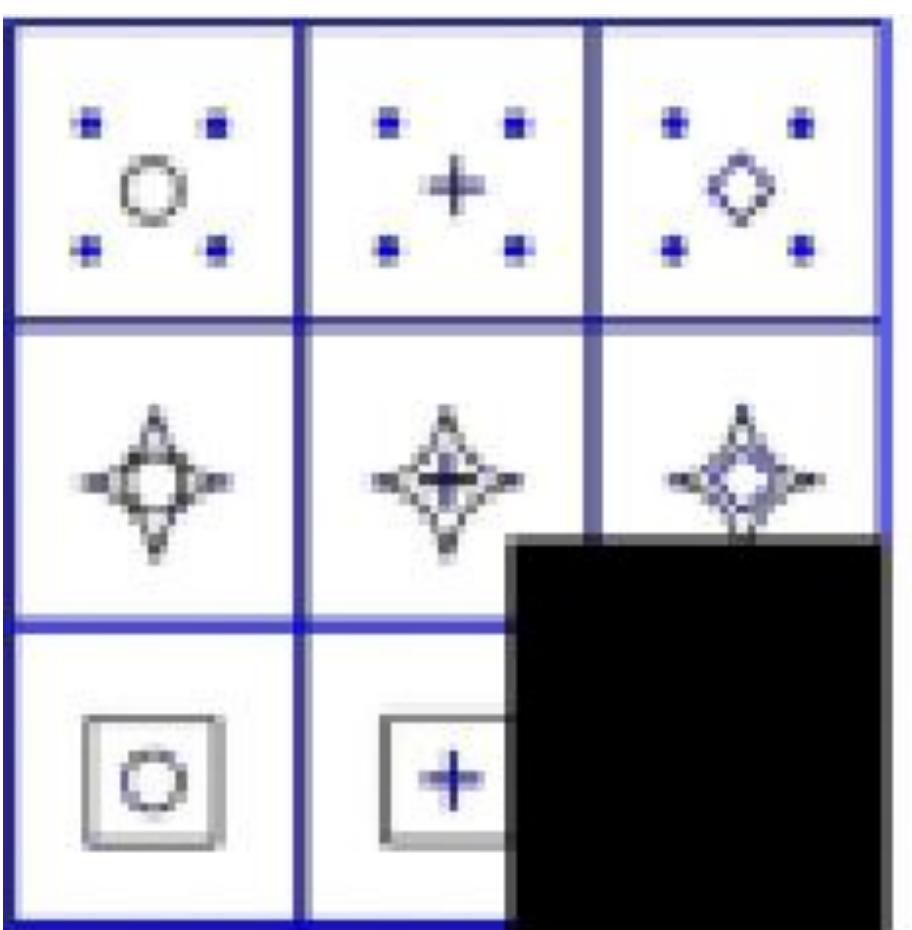


- a) b)
- c) d)
- e) f)

Raven's Progressive Test (Non-verbal IQ test)

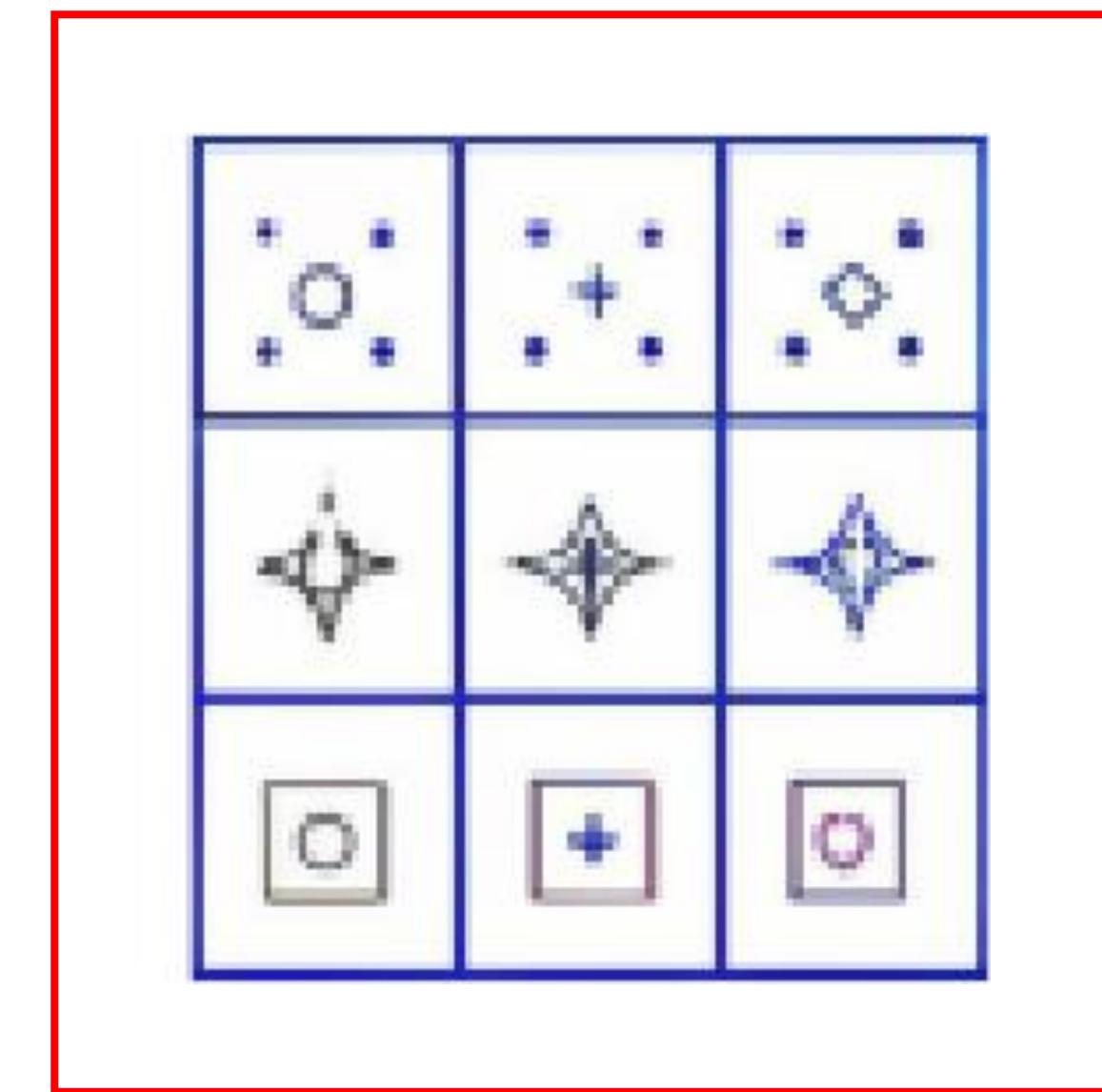
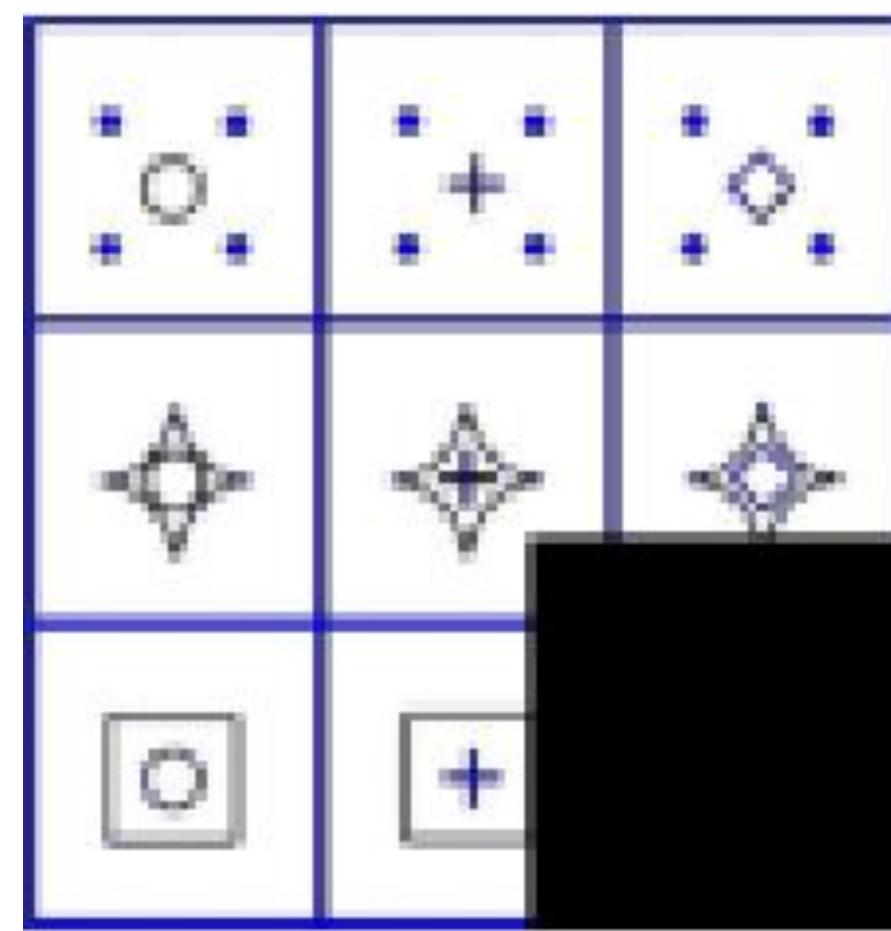


More Difficult?

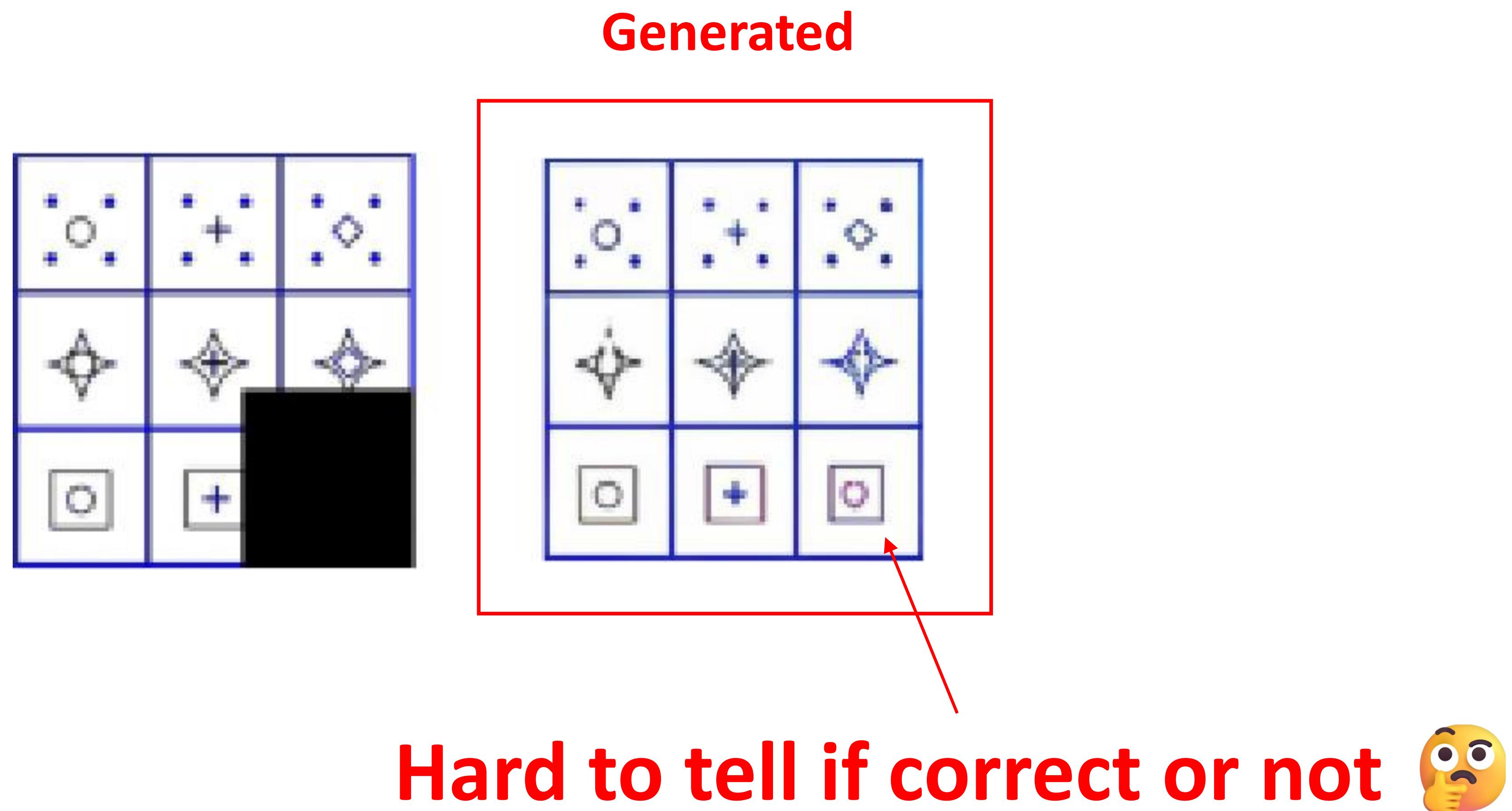


More Difficult?

Generated



More Difficult?



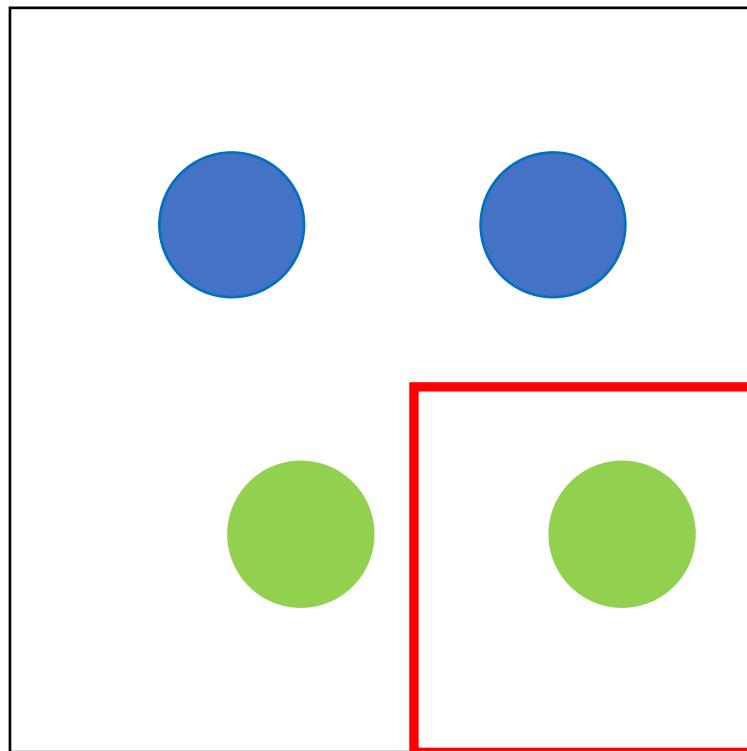
Perplexity

- Small Quantitative study (n=10):
 - perplexity analysis on classic Raven 5-way multiple-choice Matrices, choosing the answer with lowest perplexity.

Raven's Progressive Matrices	
Chance	20%
Ours	30%

Synthetic Reasoning

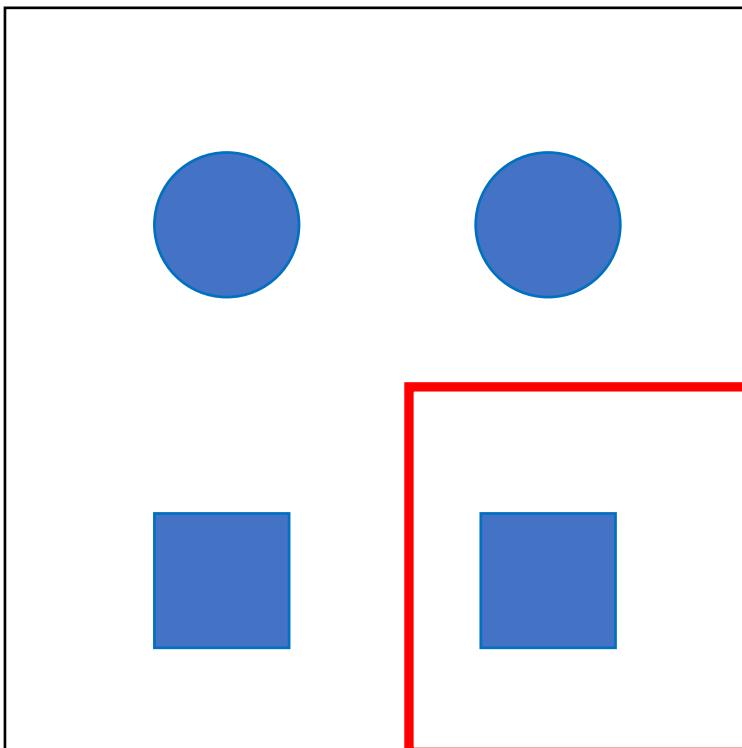
- **Color Change:** choose from 3 random generated colors.



	color
Chance	33%
Ours	42%

Synthetic Reasoning

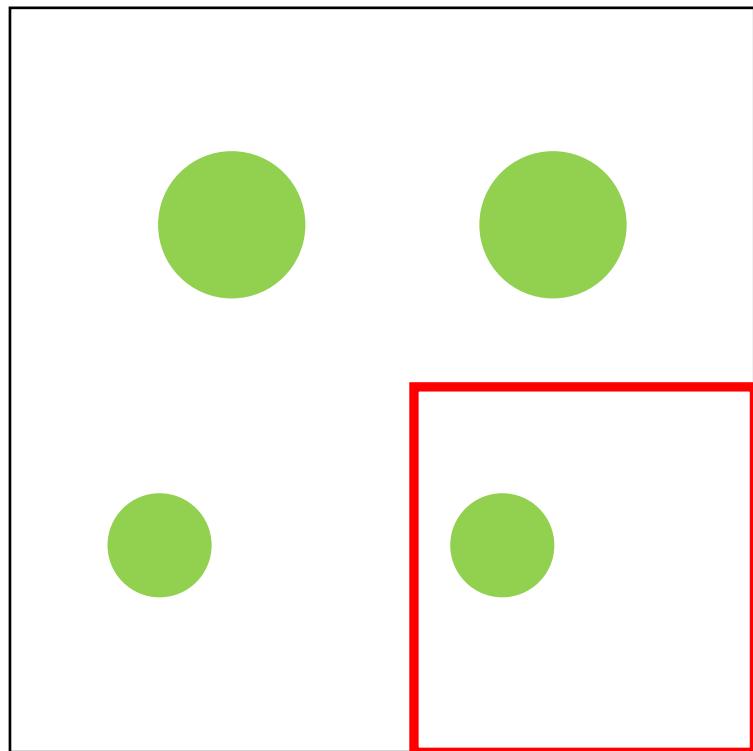
- **Shape Change:** choose from 3 random generated shapes.



	shape
Chance	33%
Ours	45%

Synthetic Reasoning

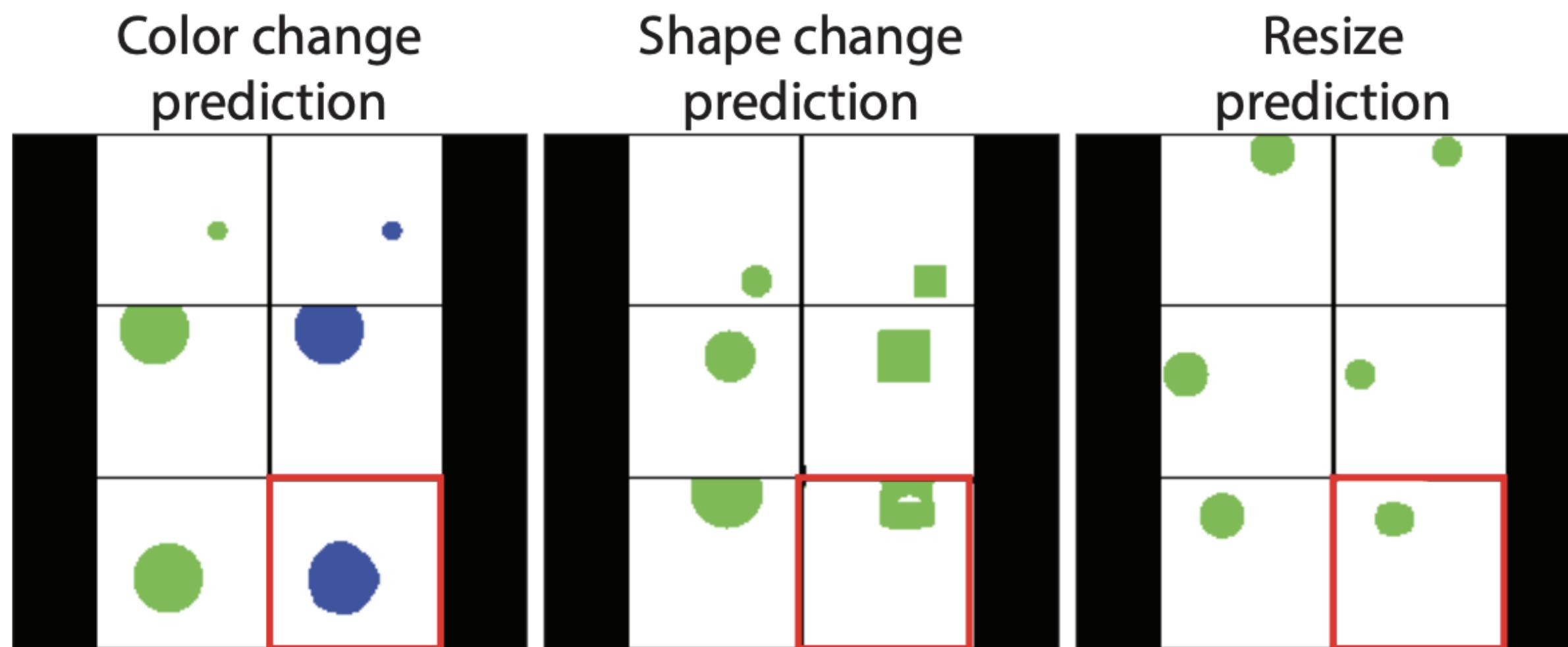
- **Size Change:** choose from 2 random generated sizes. (resolution)



	size
Chance	50%
Ours	94%

Synthetic Reasoning

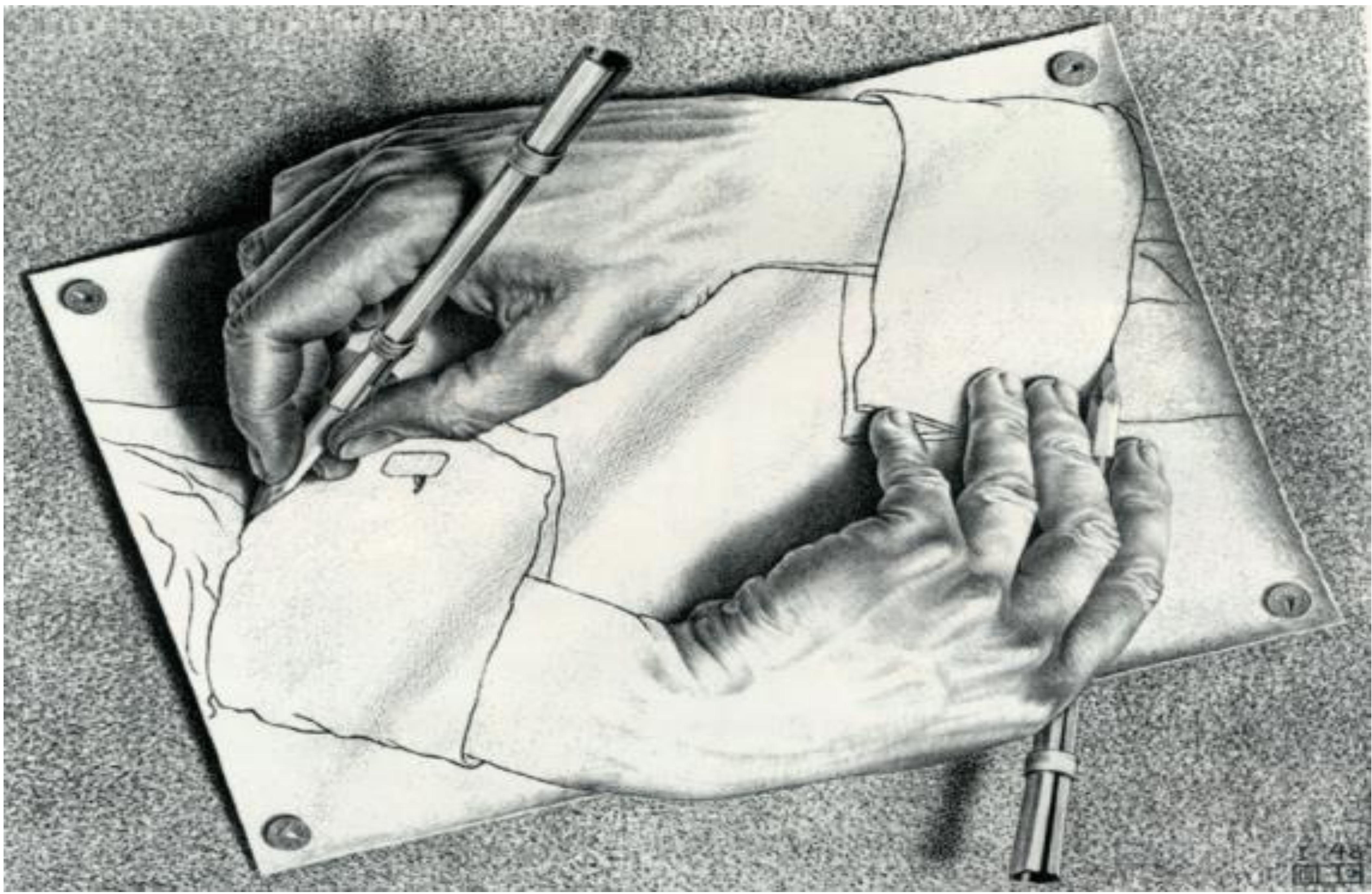
- In total 900 experiments



	color	shape	size
Chance	33%	33%	50%
Ours	42%	45%	94%

Summary

- The dream of self-supervising ourselves with **natural world data** (rather than text) is alive and well!
- But we are still at the beginning of the journey
- Plenty left to do!



王生昌
印