

Sequence Models for words and pixels



© A.A. Efros

Many slides from Steve Seitz's wonderful [5 min Lectures](#)

CS194: Intro to Computer Vision & Comp. Photography
Alexei Efros, UC Berkeley, Fall 2024

Michel Gondry train video

<http://www.youtube.com/watch?v=0S43lwBF0uM>

“Amateur” by Lasse Gjertsen

<http://www.youtube.com/watch?v=JzqumbhfxRo>

Generative AI

Generative models trained on **lots of data** have revolutionized computer science and beyond!



Text

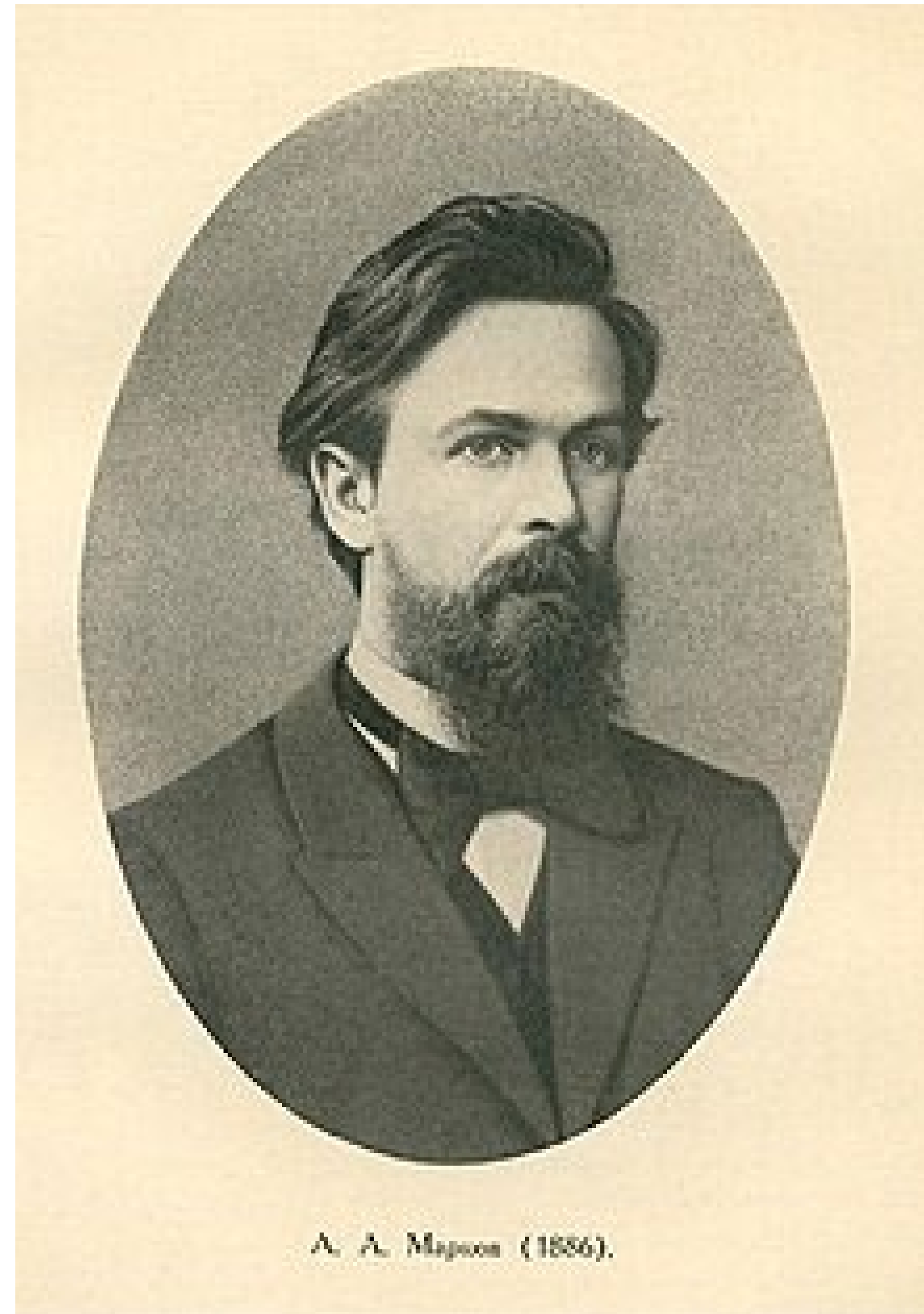


Pixels



Audio

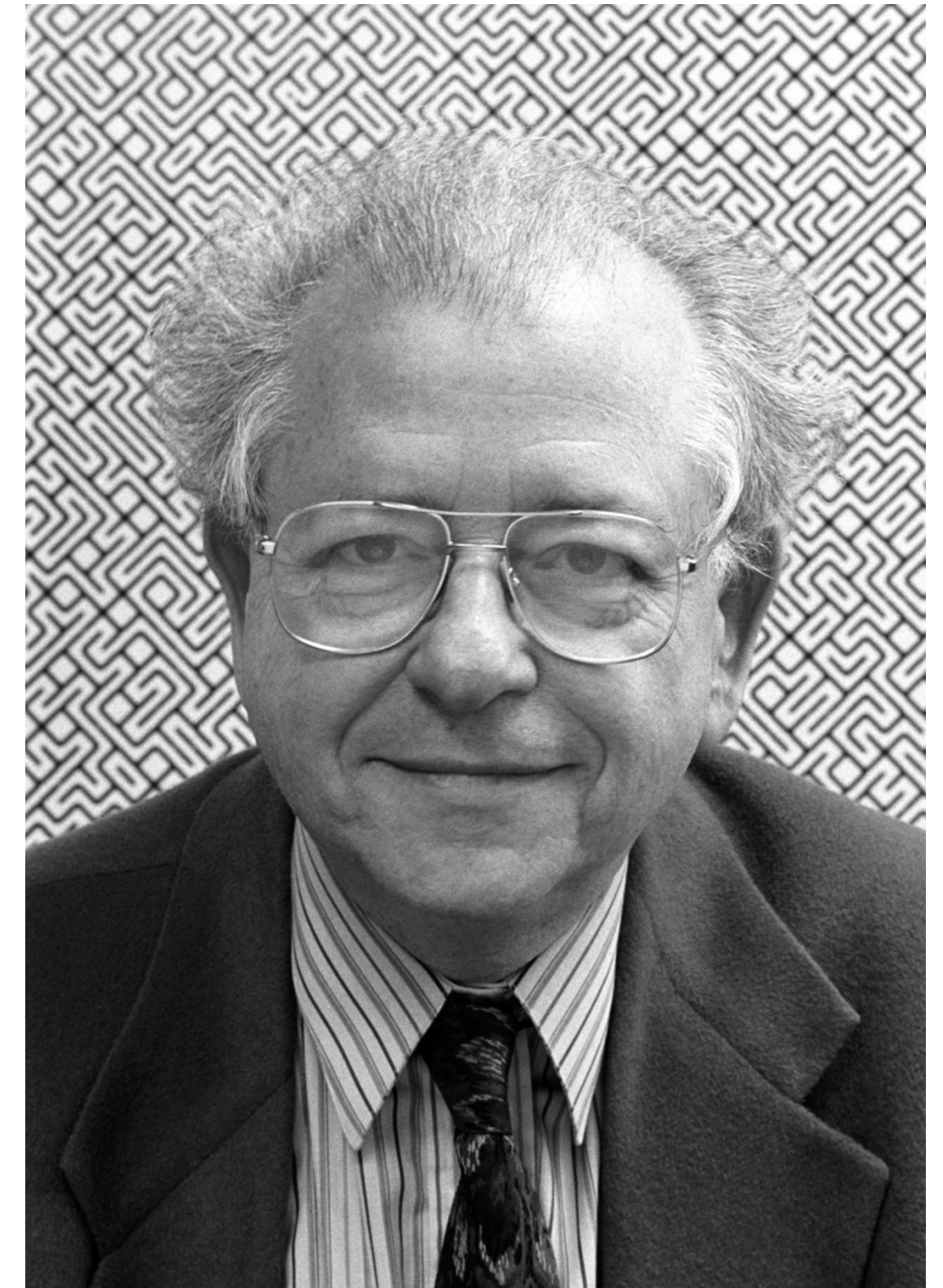
Ideas going back over a hundred years...



**Andrey Markov
(1856-1922)**



**Claude Shannon
(1916-2001)**



**Bela Julesz
(1928-2003)**

Weather Forecasting for Dummies™

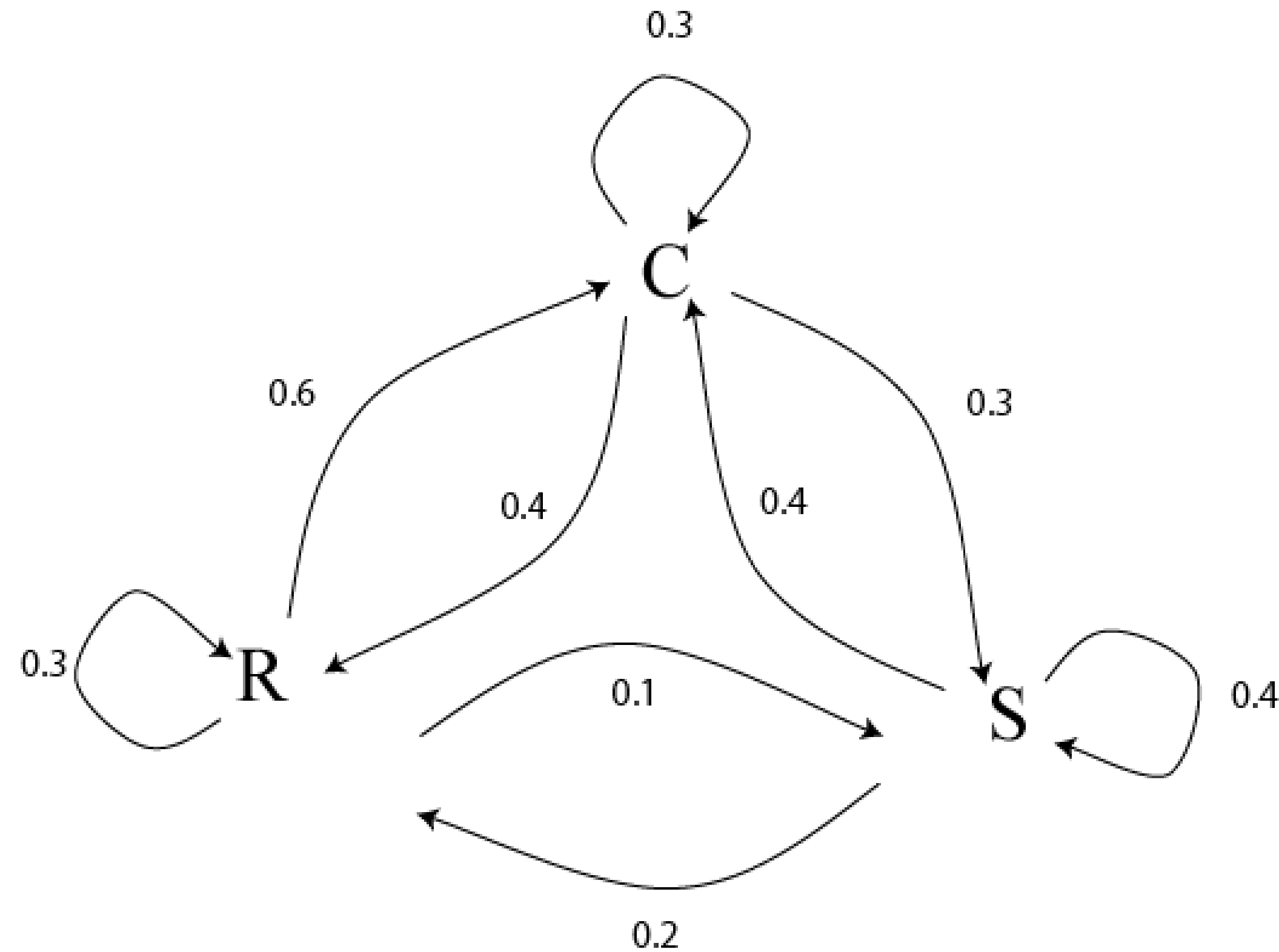
Let's predict weather:

- 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 yestarday's weather?

Now, let's apply this to text

[Markov, 1913] statistical analysis of text

[Shannon, 1948] proposed a way to generate English-looking text using N-grams:

- Assume a generalized Markov model of language
- 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
- E.g.:

WE NEED TO EAT CAKE



Mark V. Shaney (Bell Labs)

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*

~~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~~

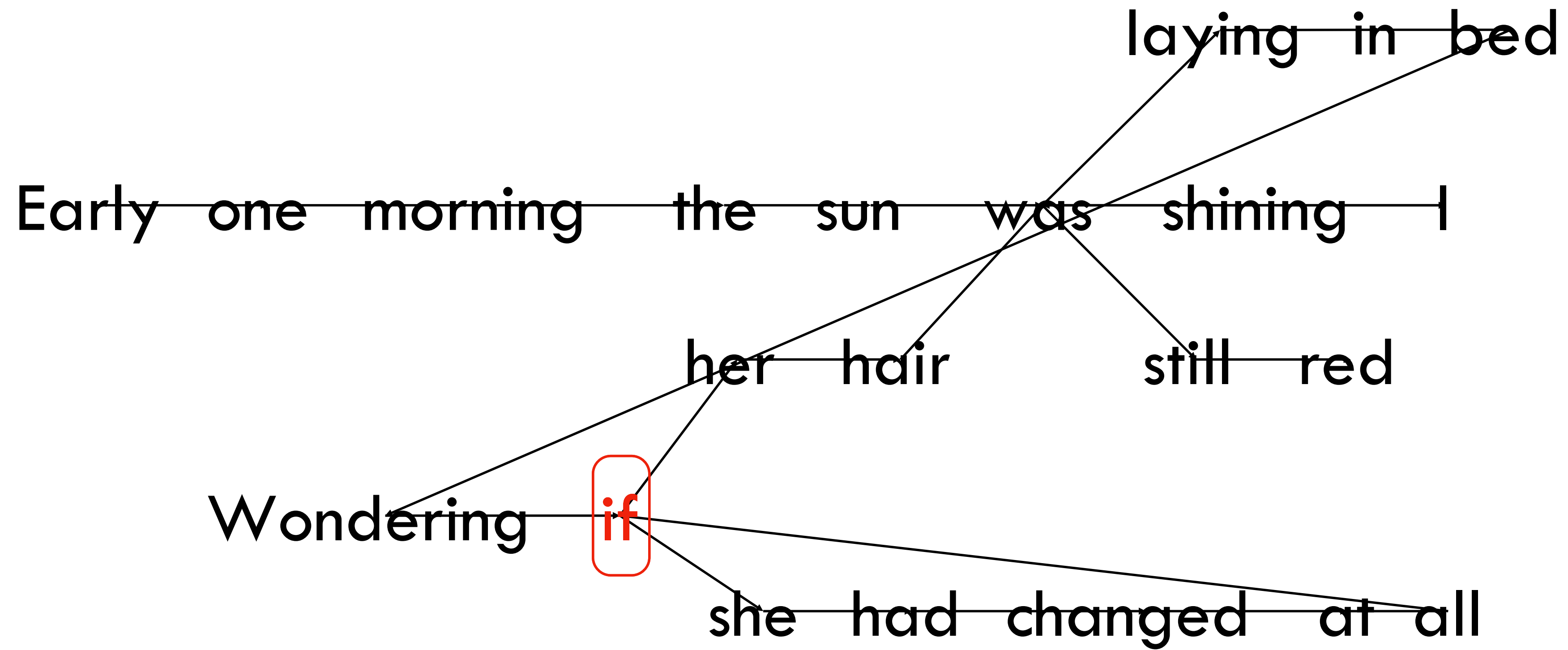
Early one morning the sun **was** shining I **was** laying in bed

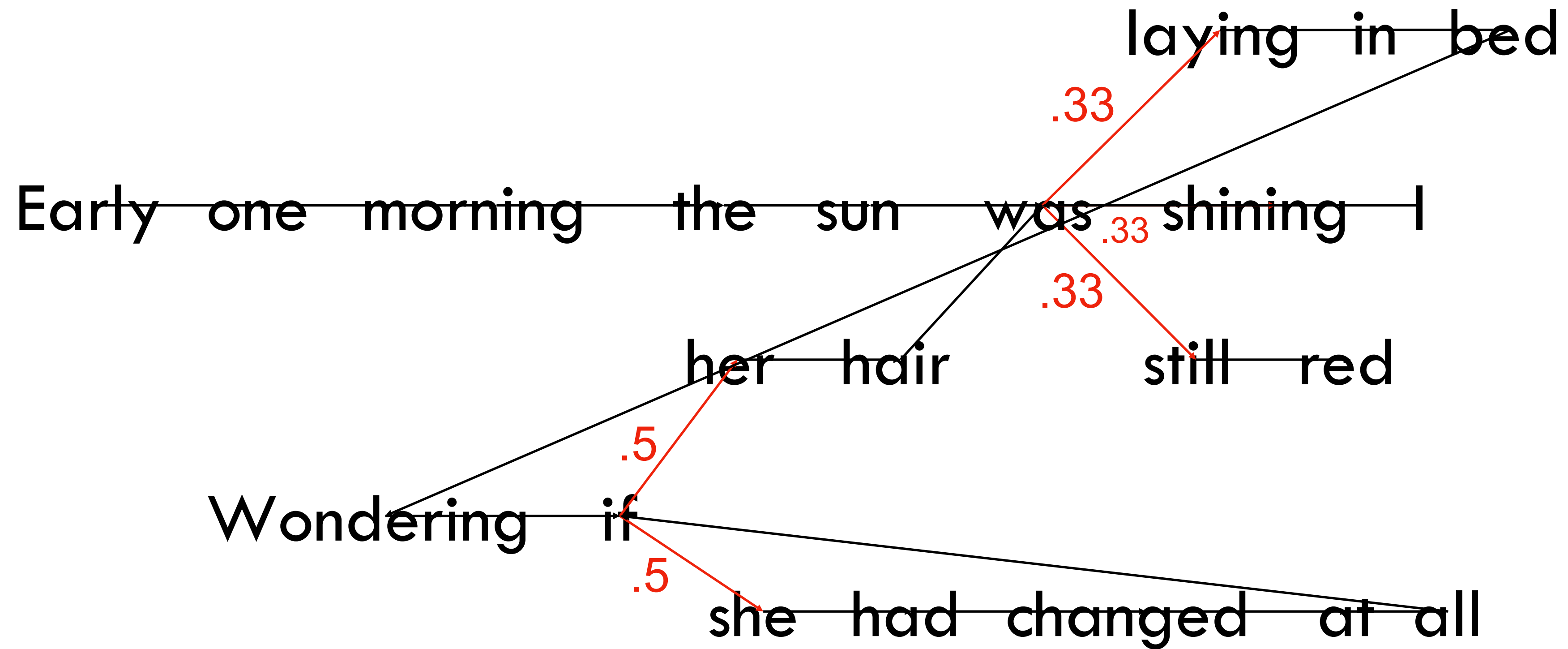
Wondering if she had changed at all if her hair **was** still red

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

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

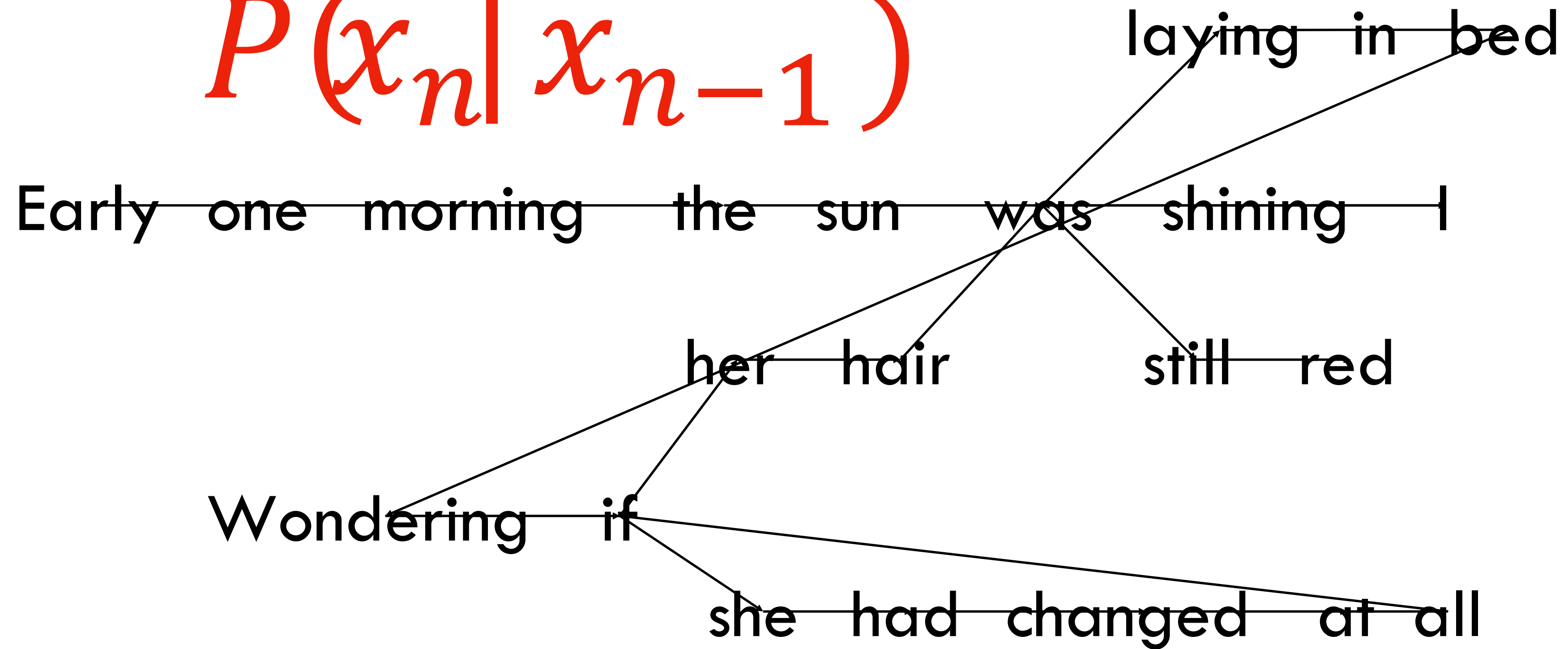
```
graph TD; was[was] --- laying[laying in bed]; was --- hair[her hair]; if1[if] --- still[still red]; if1 --- at[at all];
```





Language Model

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



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

Early one \longrightarrow morning

one morning \longrightarrow the

morning the \longrightarrow sun

the sun \longrightarrow was

sun was \longrightarrow shining

was shining \longrightarrow I

shining I \longrightarrow was

I was \longrightarrow laying

...

Video Textures

Arno Schödl

Richard Szeliski

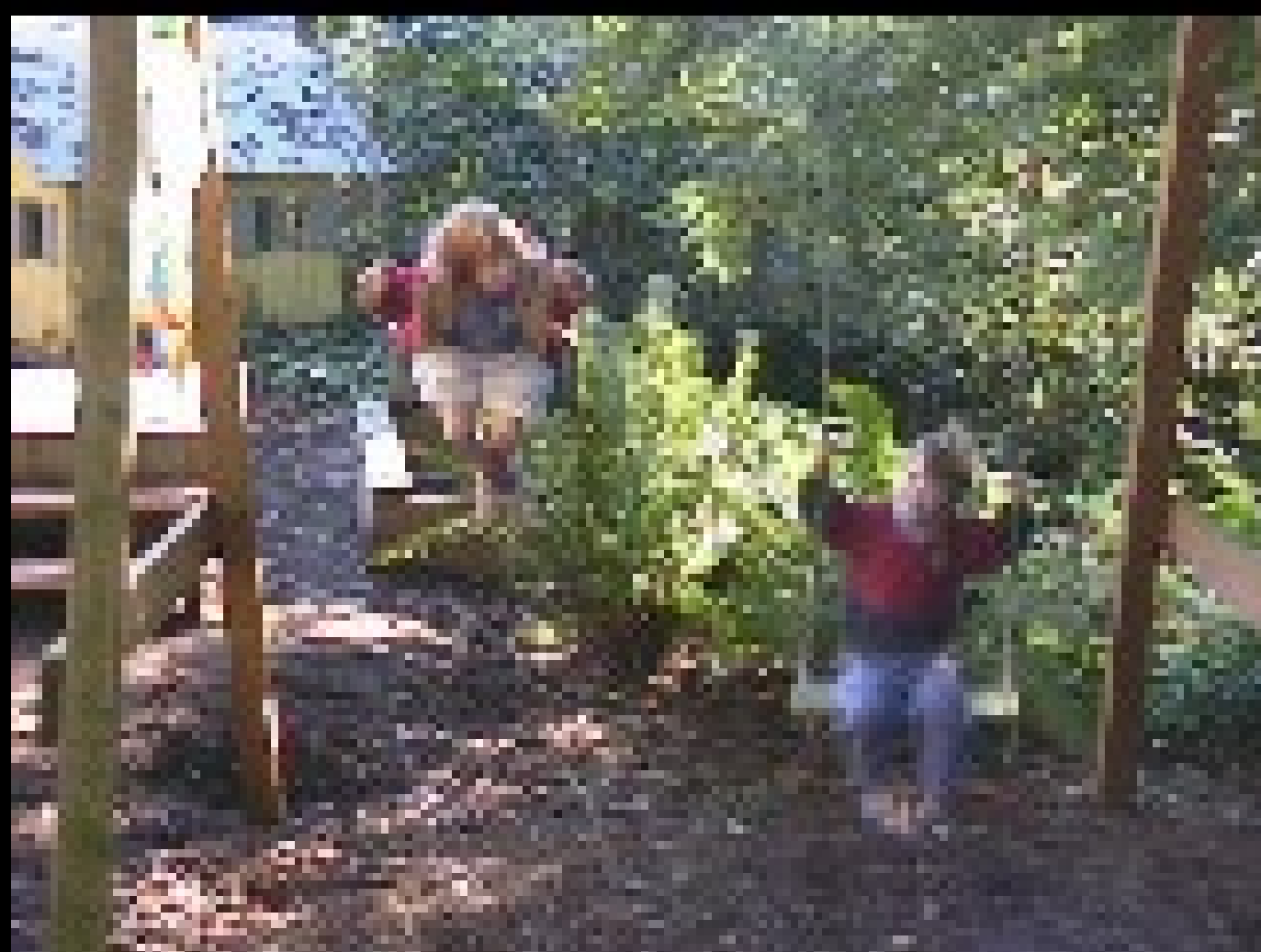
David Salesin

Irfan Essa

Microsoft Research, Georgia Tech

SIGGRAPH 2000

Still photos



Video clips



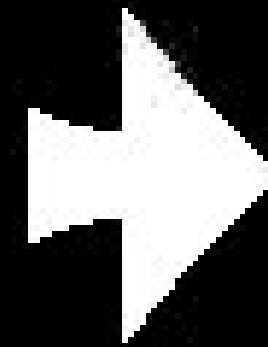
Video textures



Problem statement



video clip



video texture

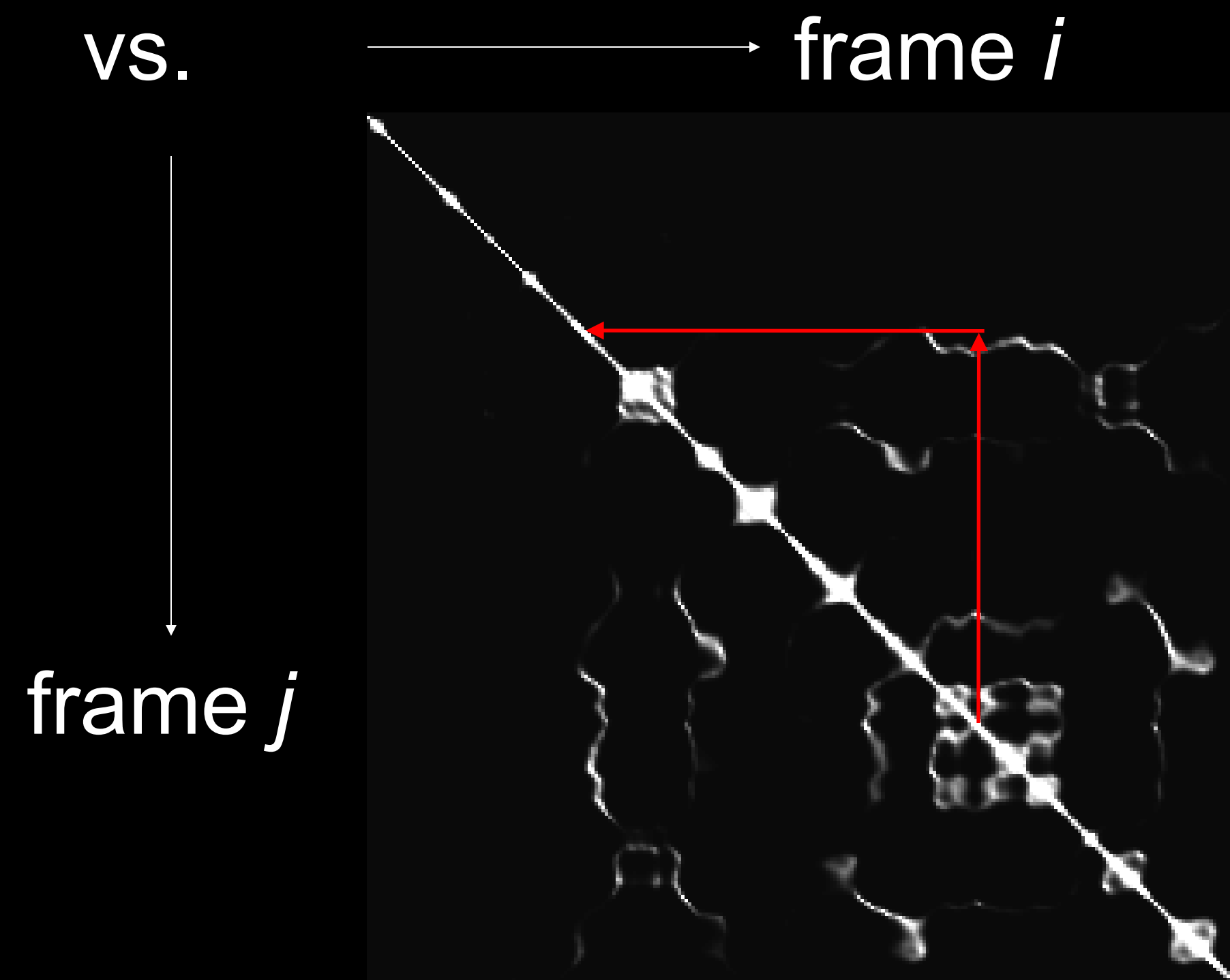
Our approach



- How do we find good transitions?

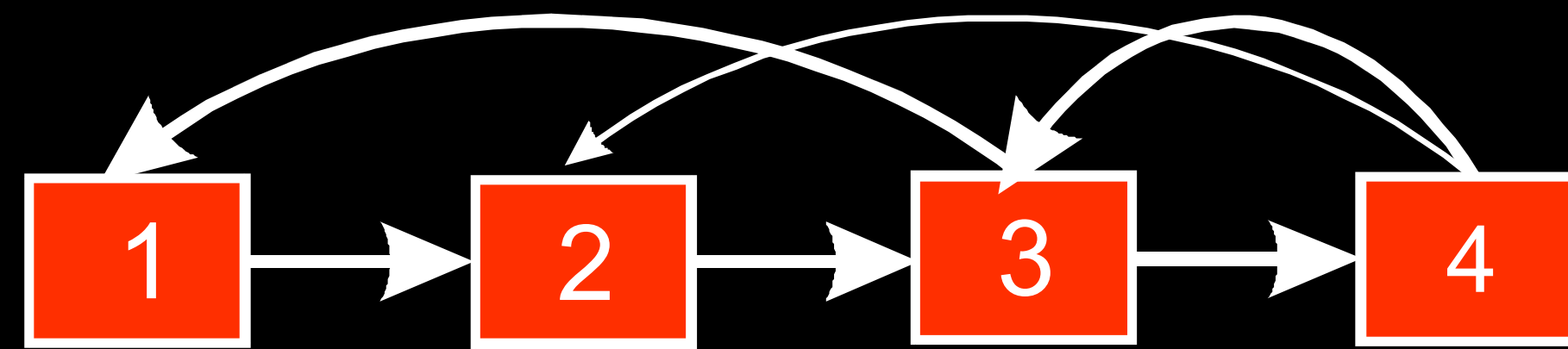
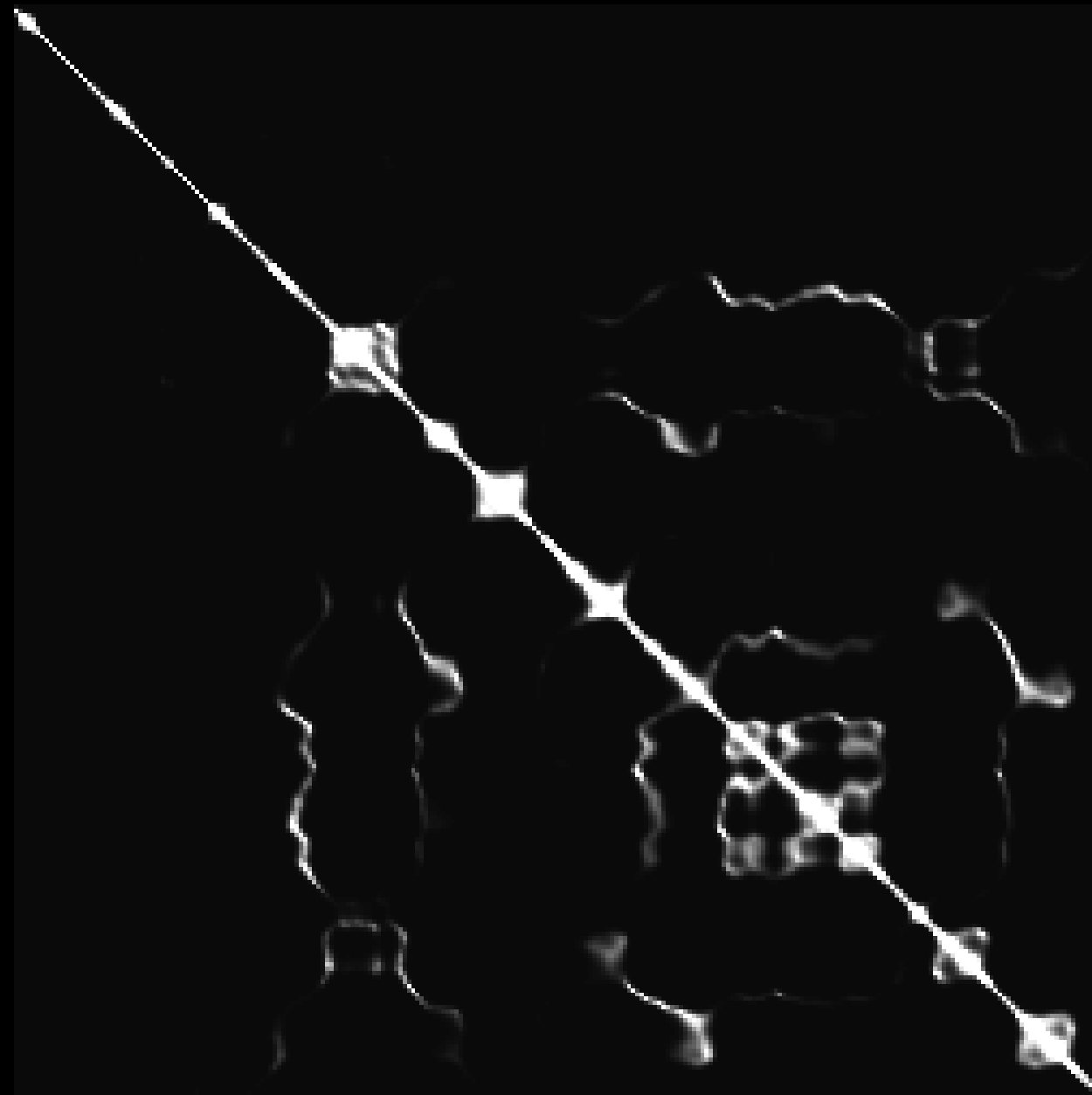
Finding good transitions

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



Similar frames make good transitions

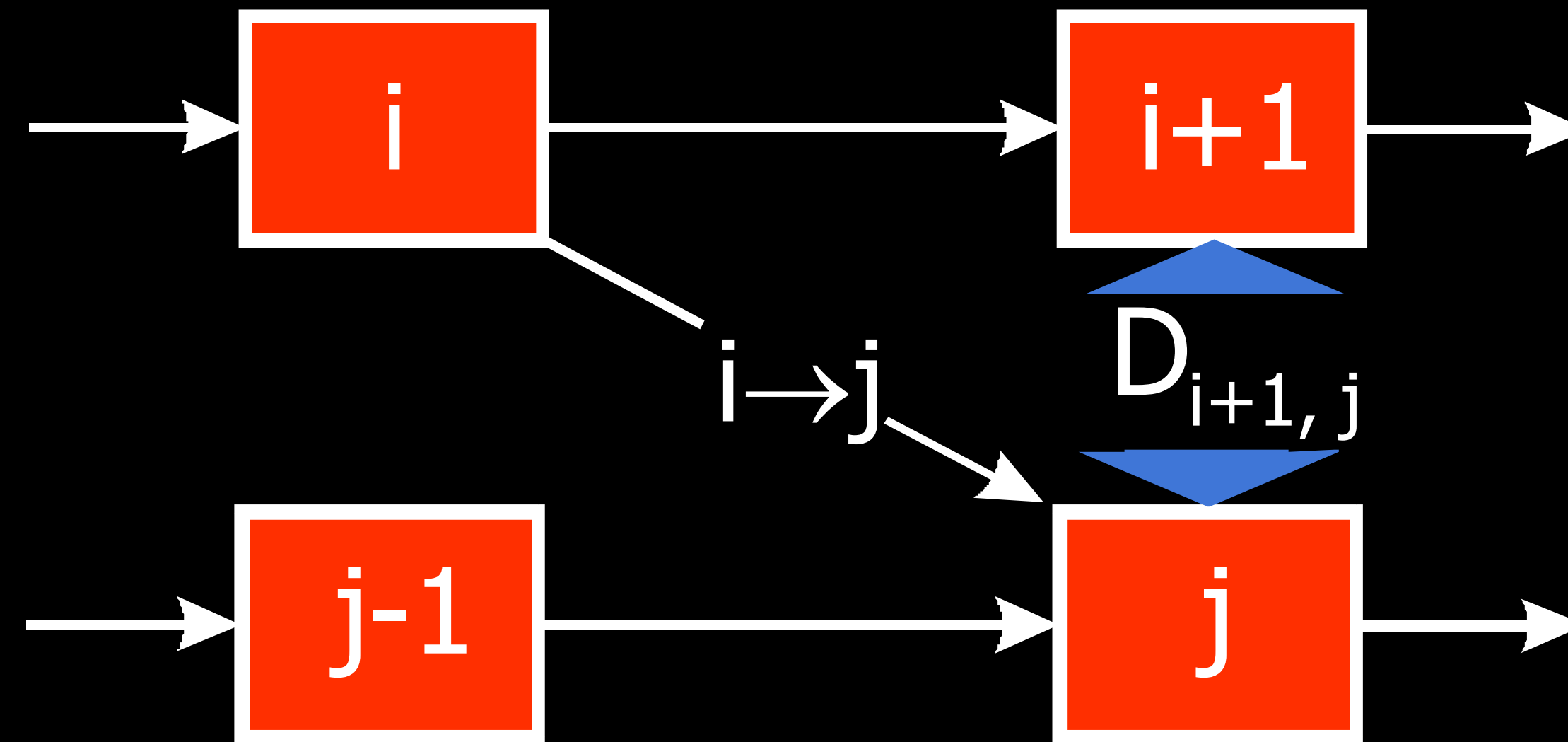
Markov chain representation



Similar frames make good transitions

Transition costs

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



Transition probabilities

- Probability for transition $P_{i \rightarrow j}$ inversely related to cost:

- $P_{i \rightarrow j} \sim \exp(-C_{i \rightarrow j} / \sigma^2)$



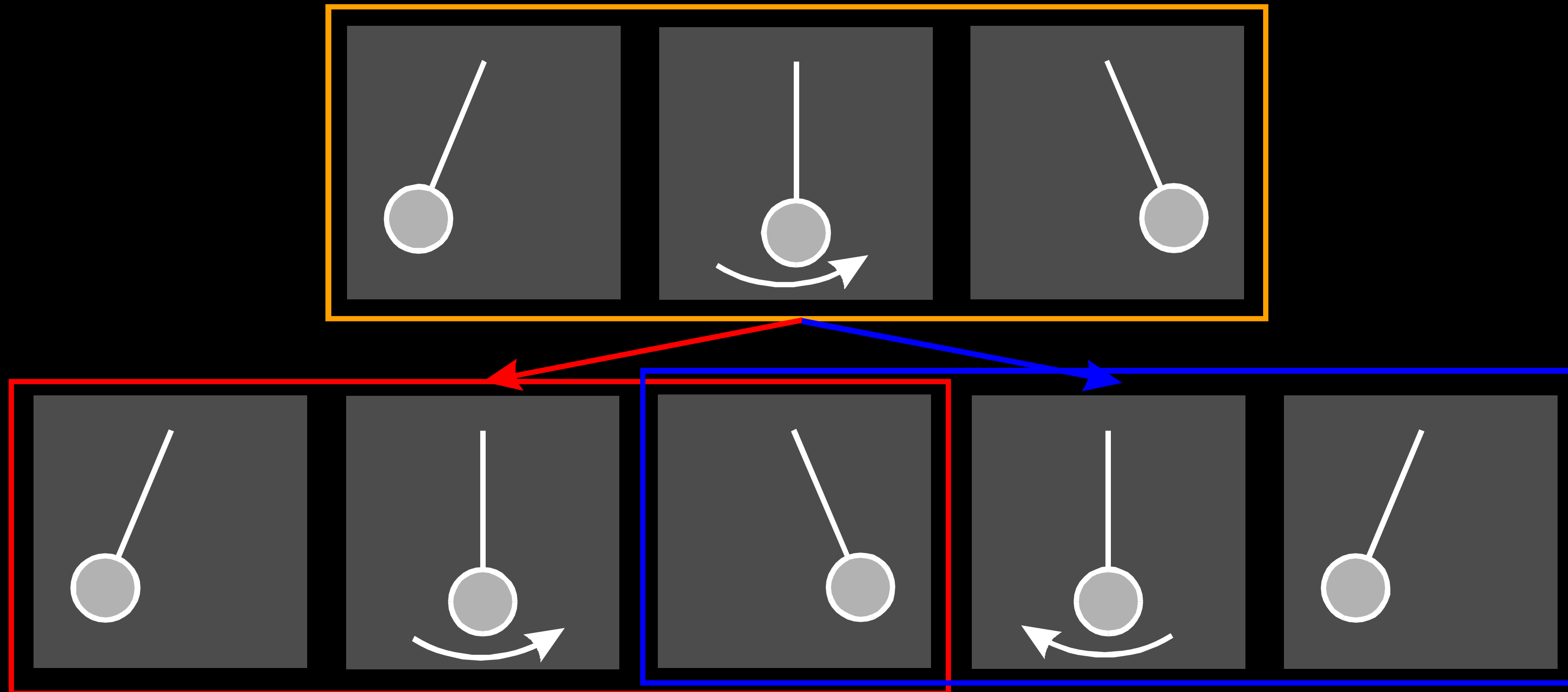
high σ

low σ

Preserving dynamics



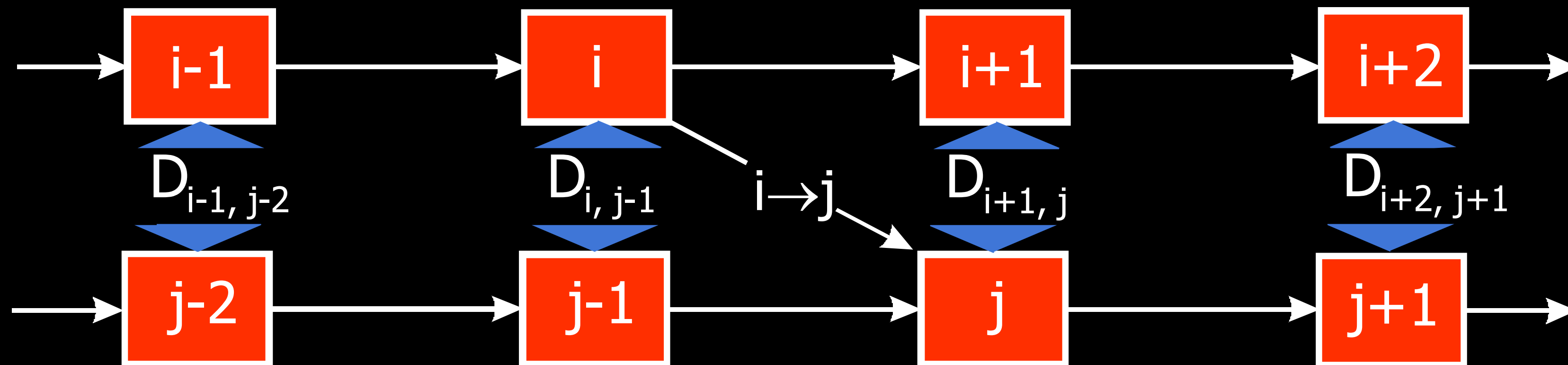
Preserving dynamics



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 $i \rightarrow j$

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



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

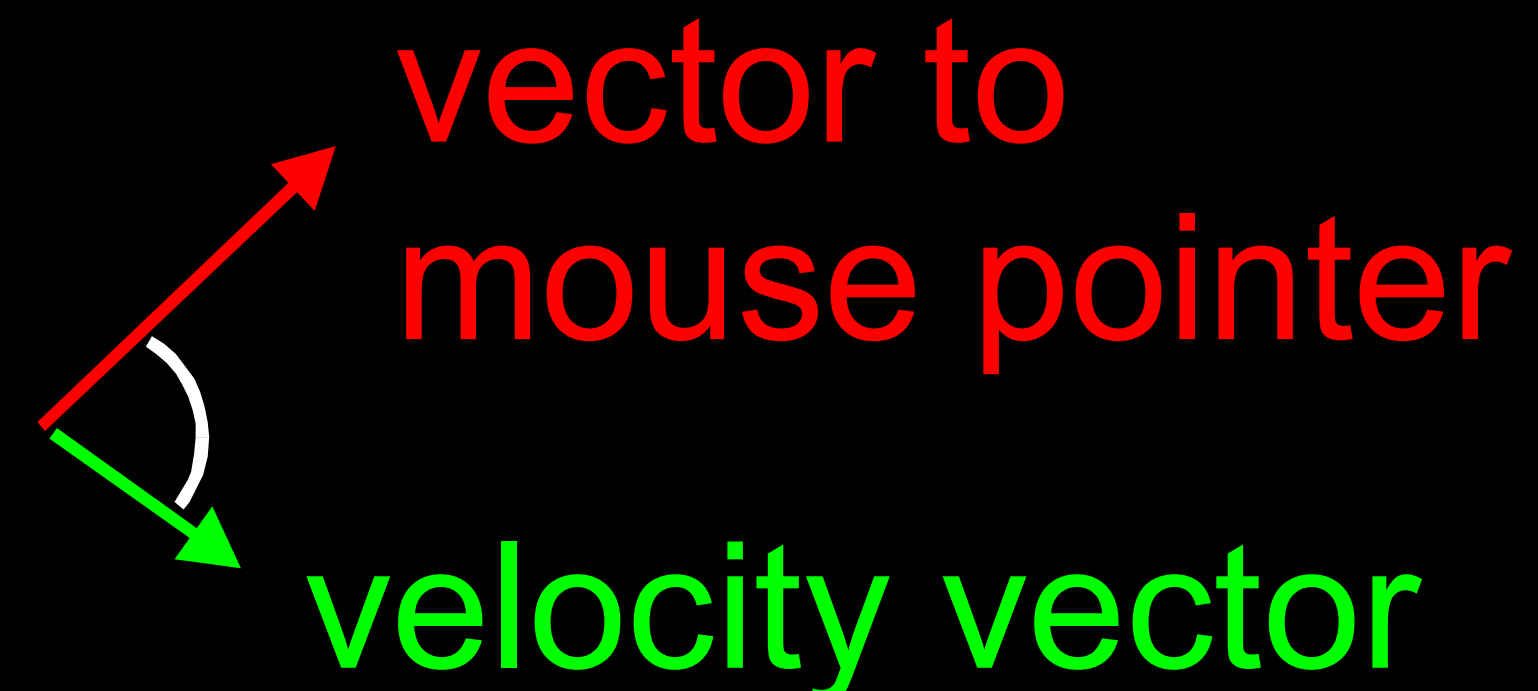


blue screen matting
and velocity estimation



Video sprite control

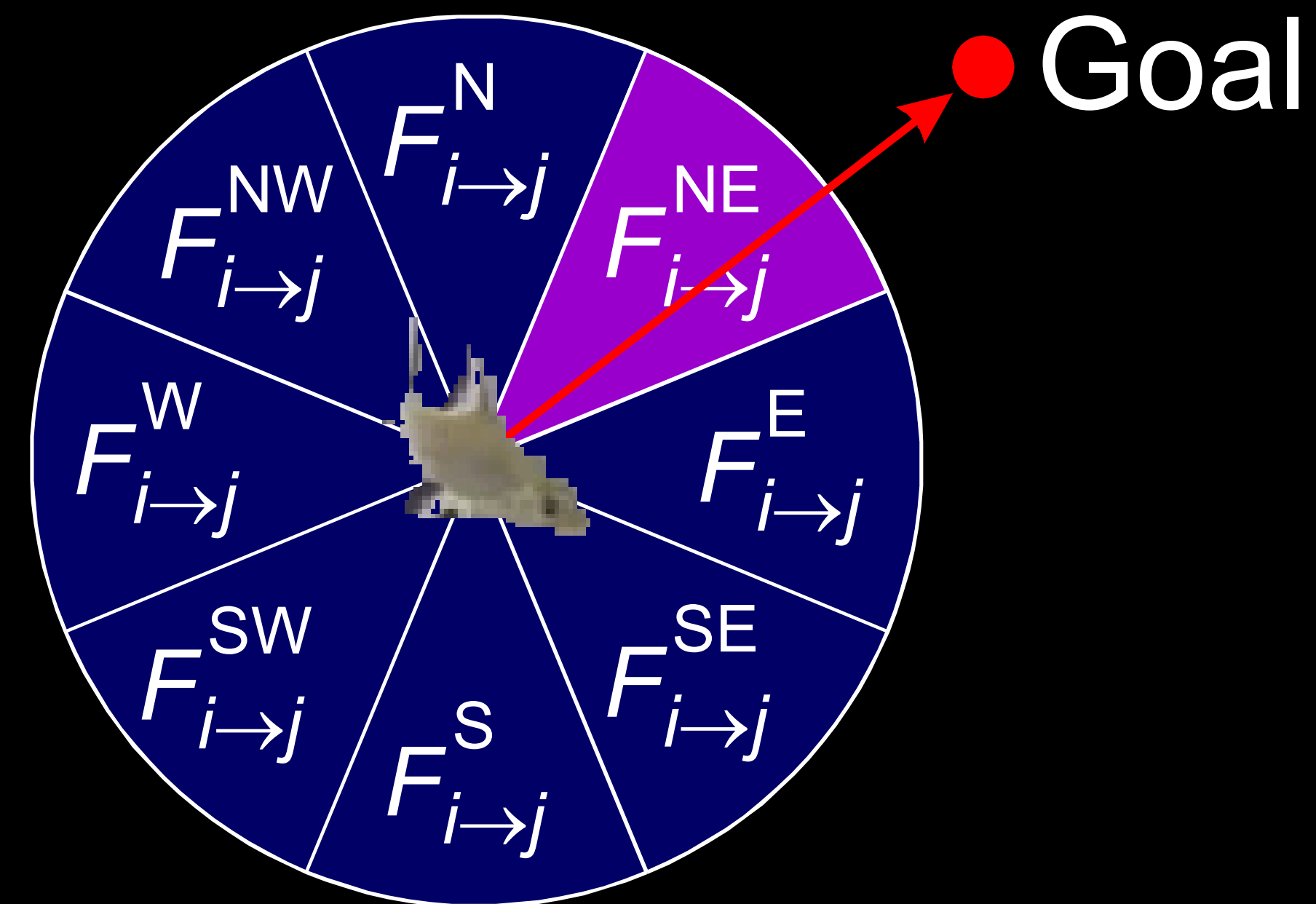
- Augmented transition cost:

$$C_{i \rightarrow j}^{\text{Animation}} = \alpha \underbrace{C_{i \rightarrow j}}_{\text{Similarity term}} + \beta \underbrace{\text{angle}}_{\text{Control term}}$$


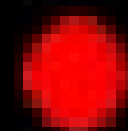
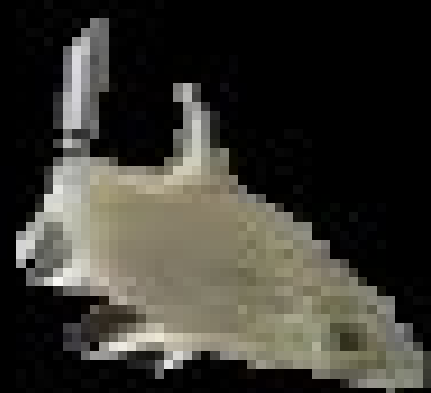
The diagram illustrates the 'Control term' in the equation. It shows two vectors originating from a common point: a red vector pointing towards the upper right, labeled 'vector to mouse pointer', and a green vector pointing towards the lower right, labeled 'velocity vector'. An arc between the two vectors indicates the angle being measured.

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



Texture

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



radishes



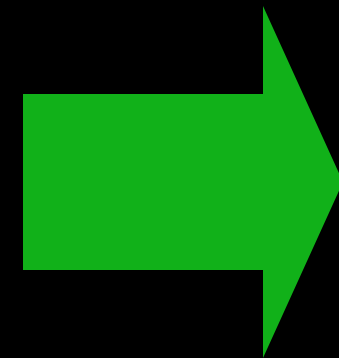
rocks



yogurt

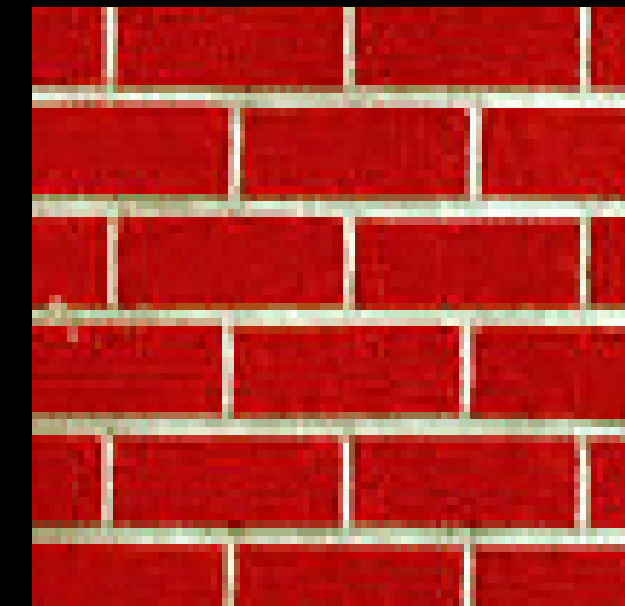
Texture Synthesis

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

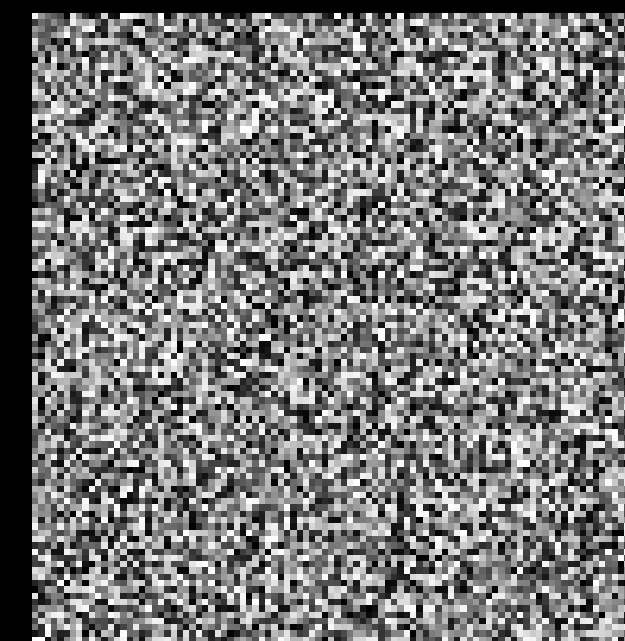


The Challenge

- Need to model the whole spectrum: from repeated to stochastic texture



repeated

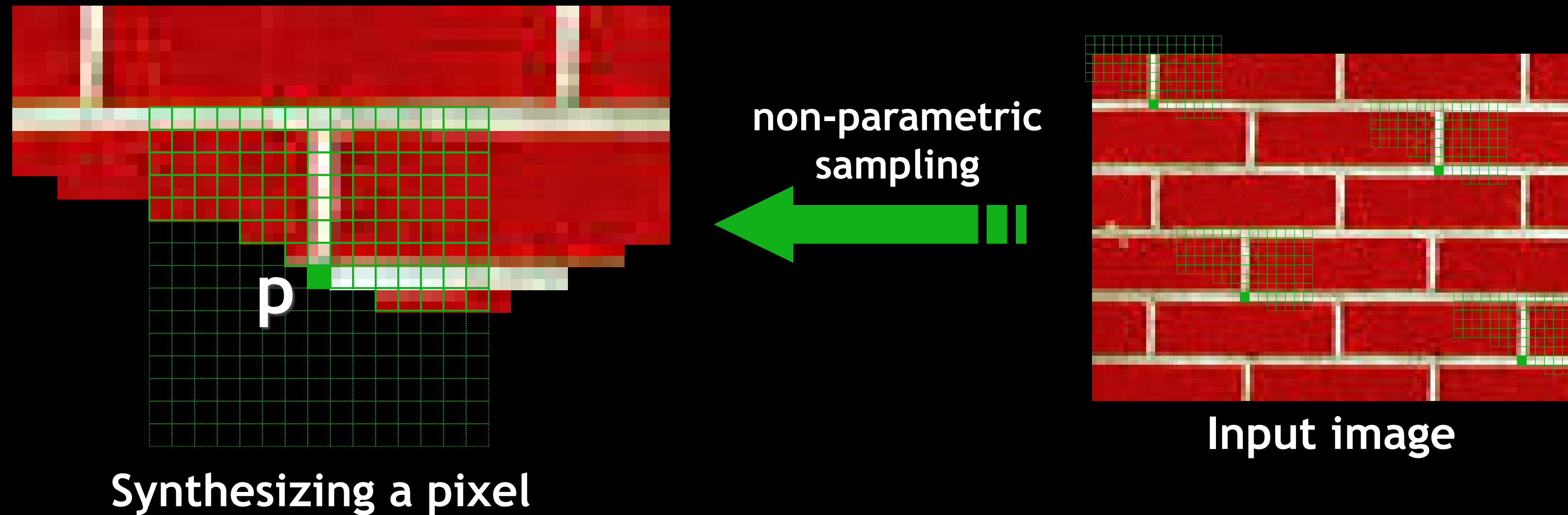


stochastic



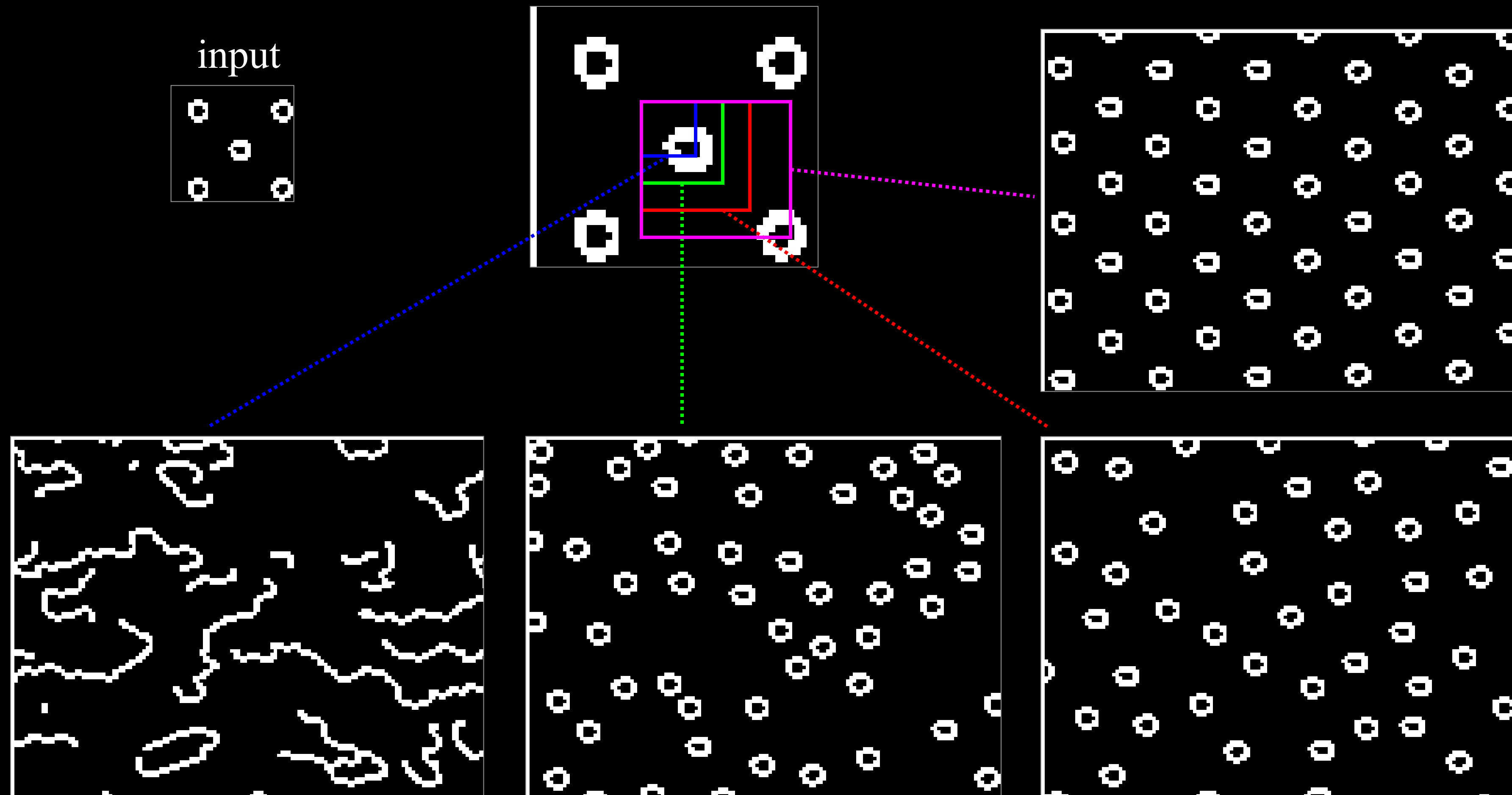
Both?

Efros & Leung Algorithm (ICCV 1999)

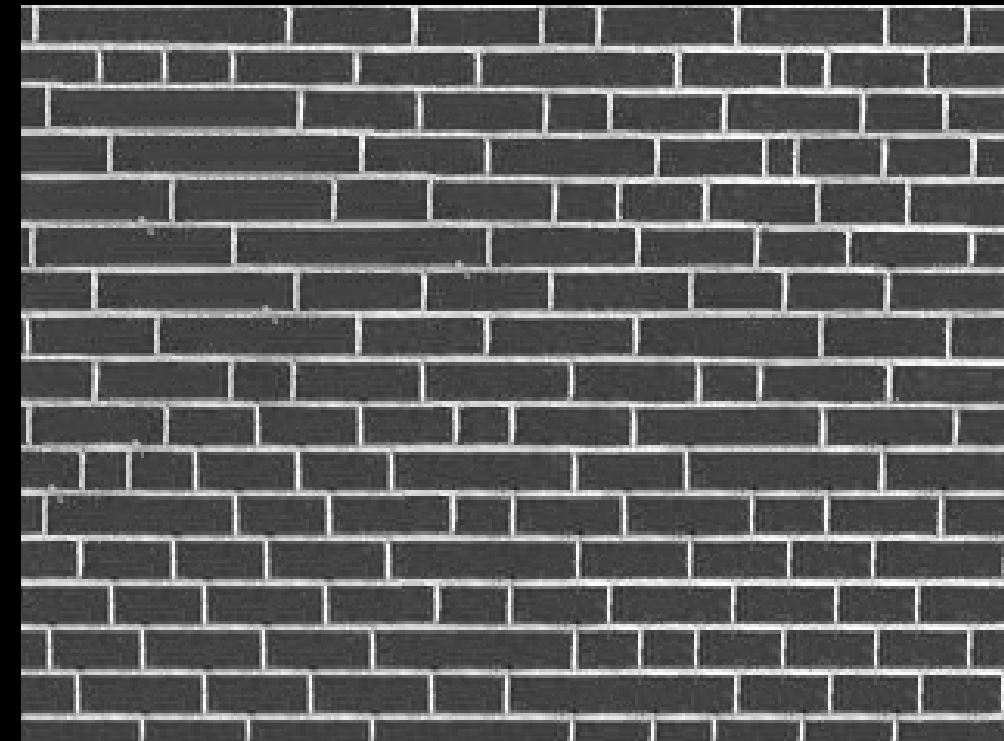
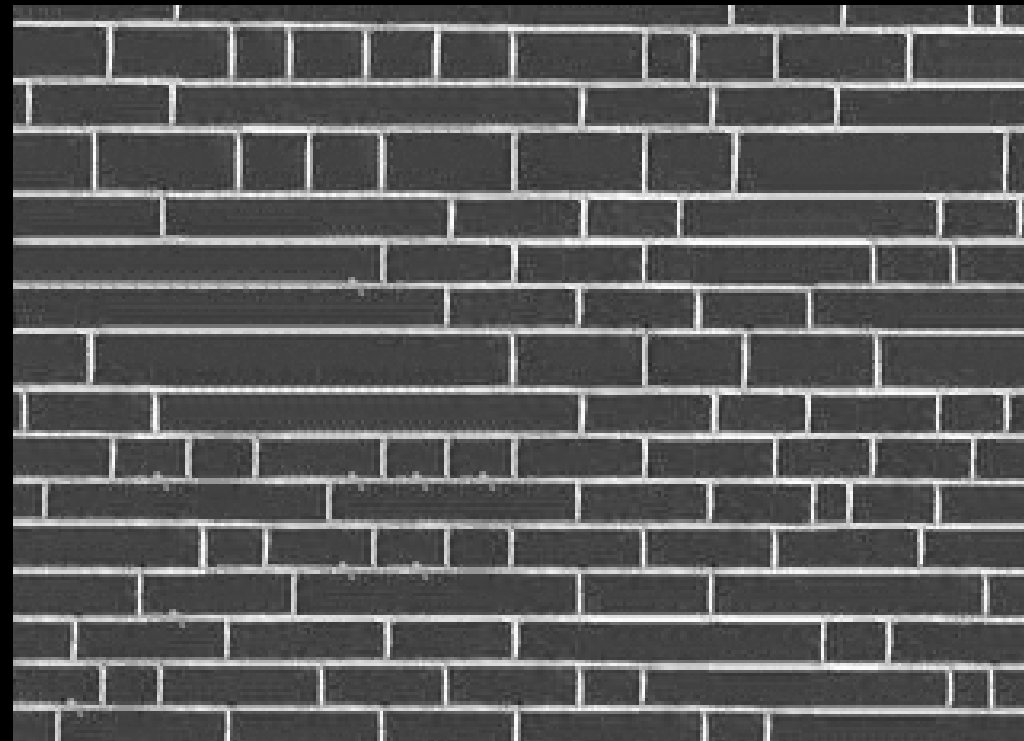
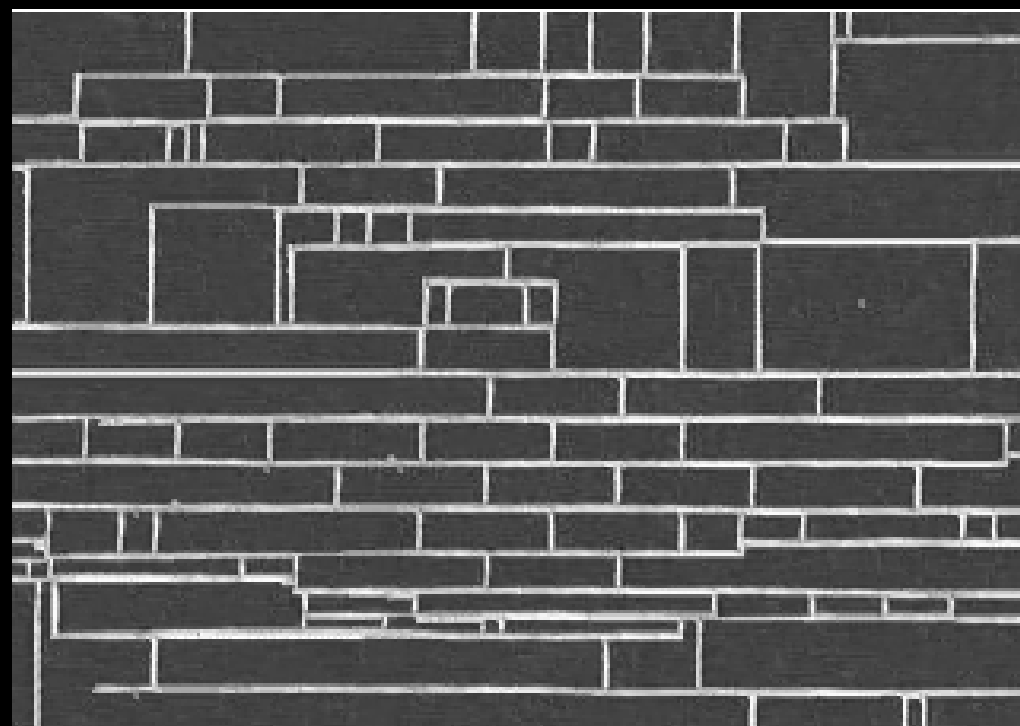
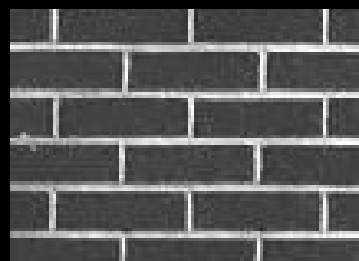
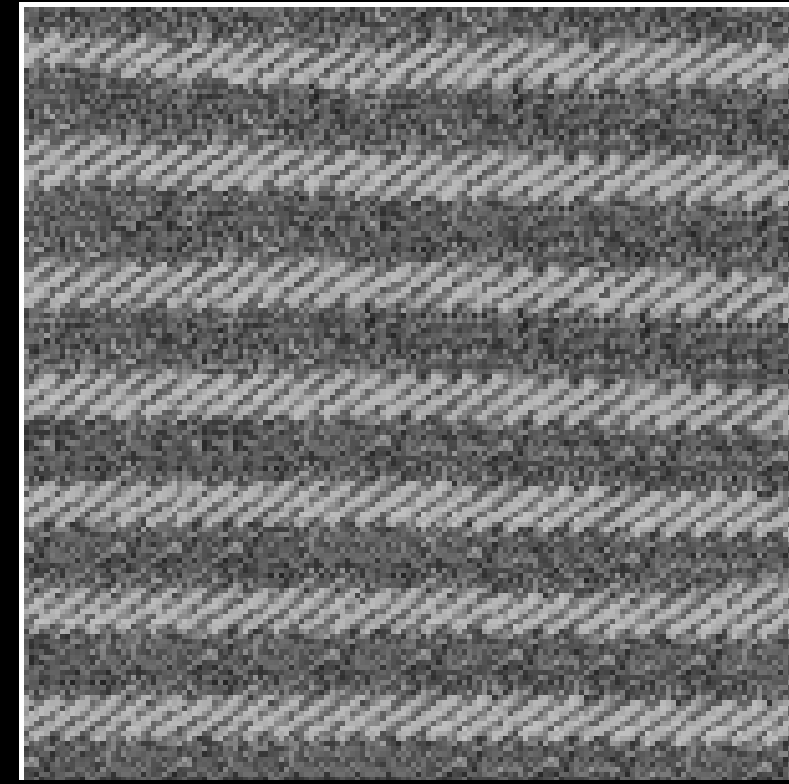
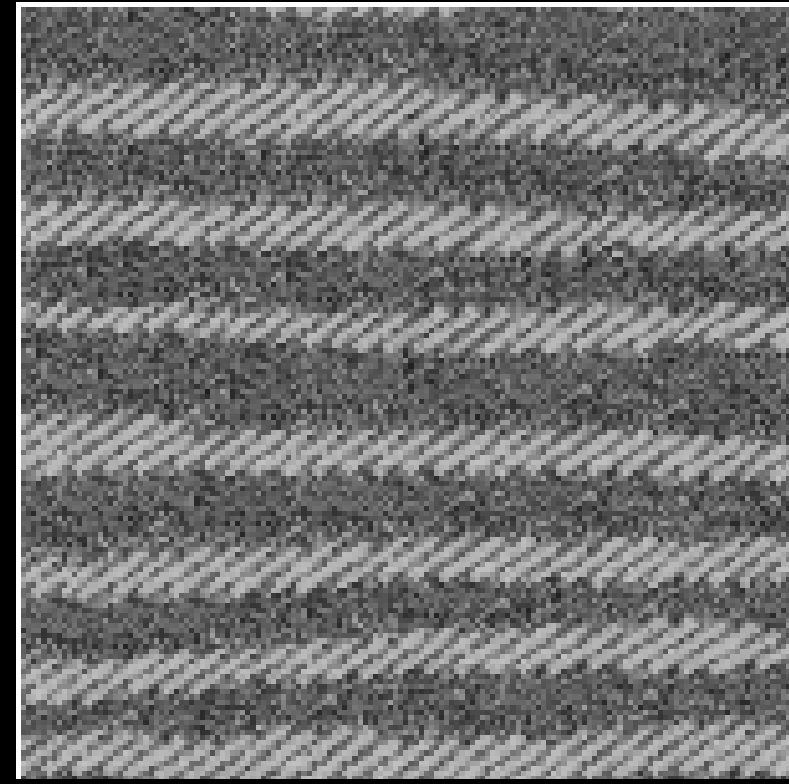
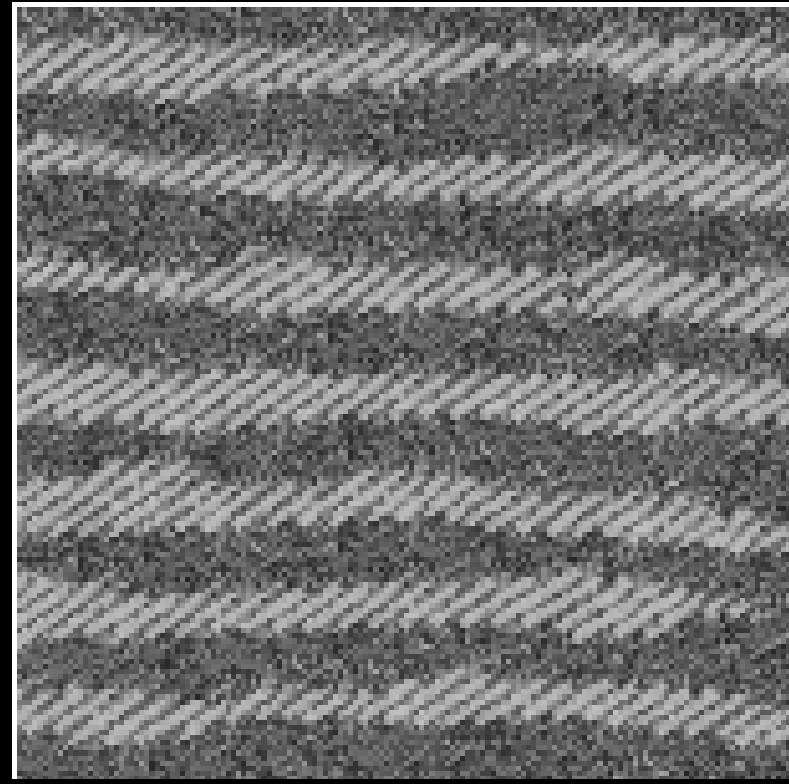
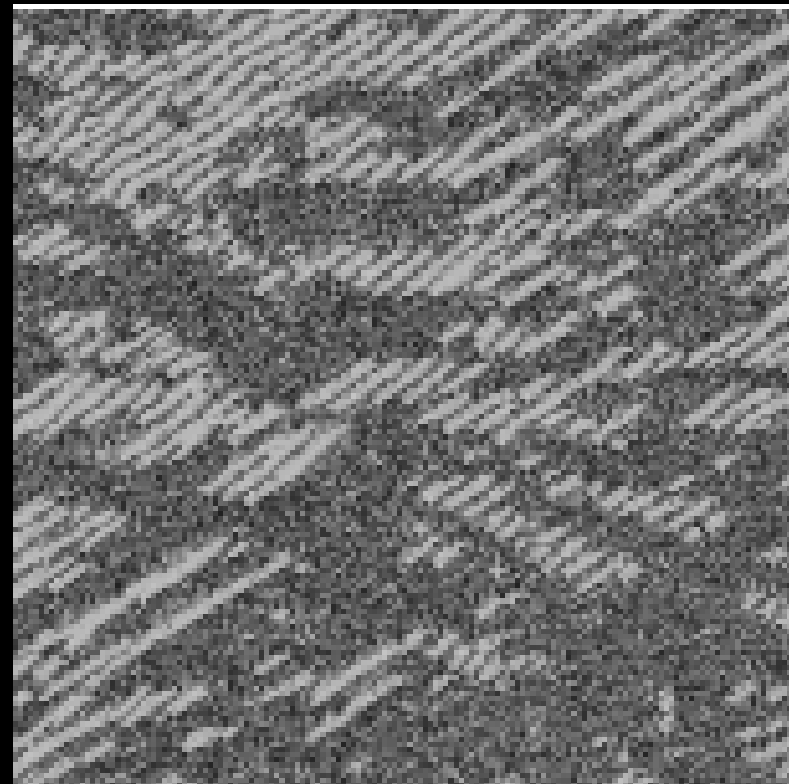
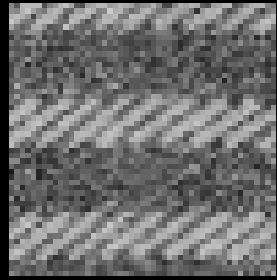


- Assuming Markov property, compute $P(\mathbf{p}|\mathbf{N}(\mathbf{p}))$
 - Building explicit probability tables infeasible
 - Instead, we *search the input image* for all similar neighborhoods — that's our pdf for \mathbf{p}
 - To sample from this pdf, just pick one match at random

Neighborhood Window



Varying Window Size

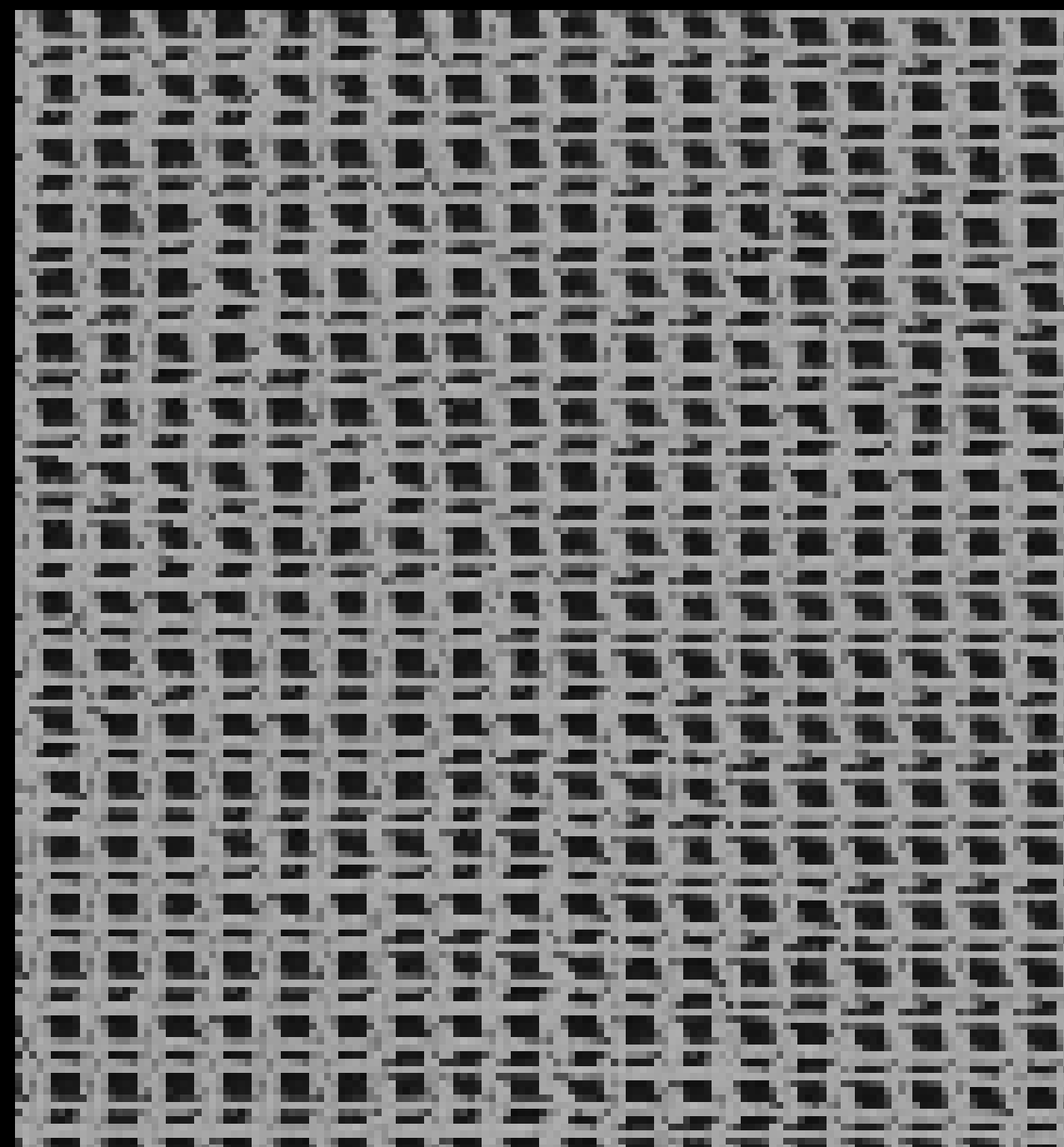
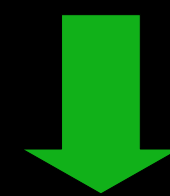
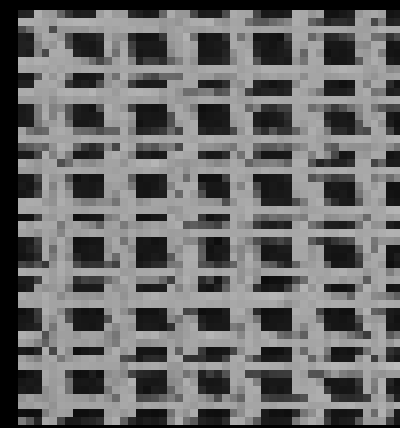


Increasing window size

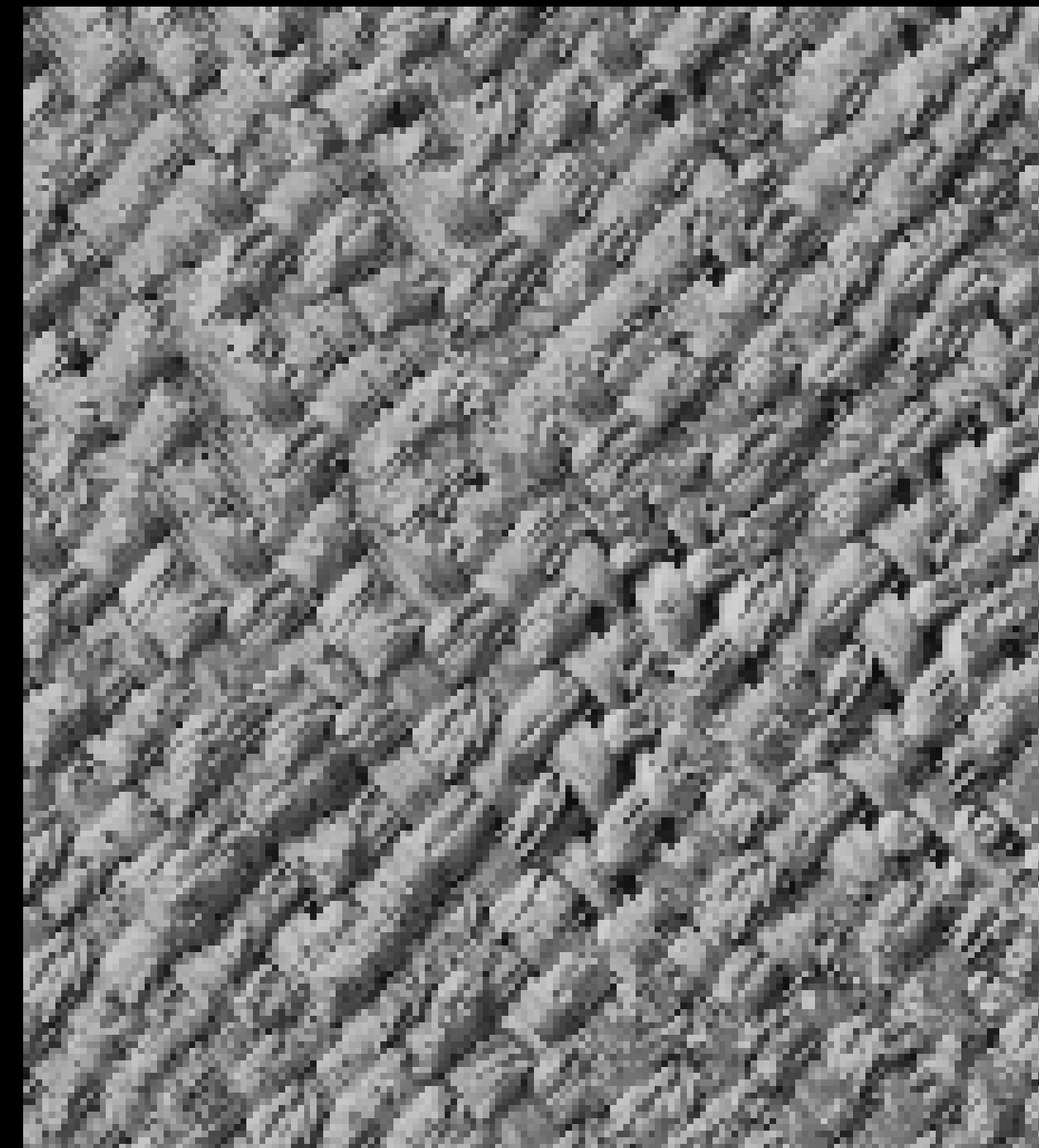
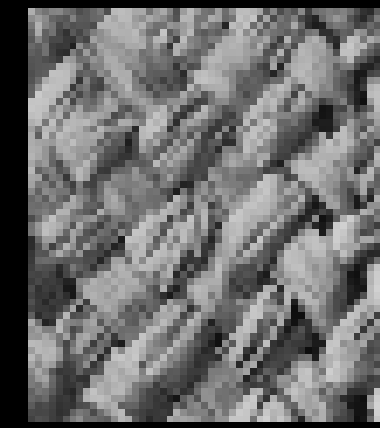


Synthesis Results

french canvas

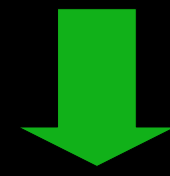


rafia weave

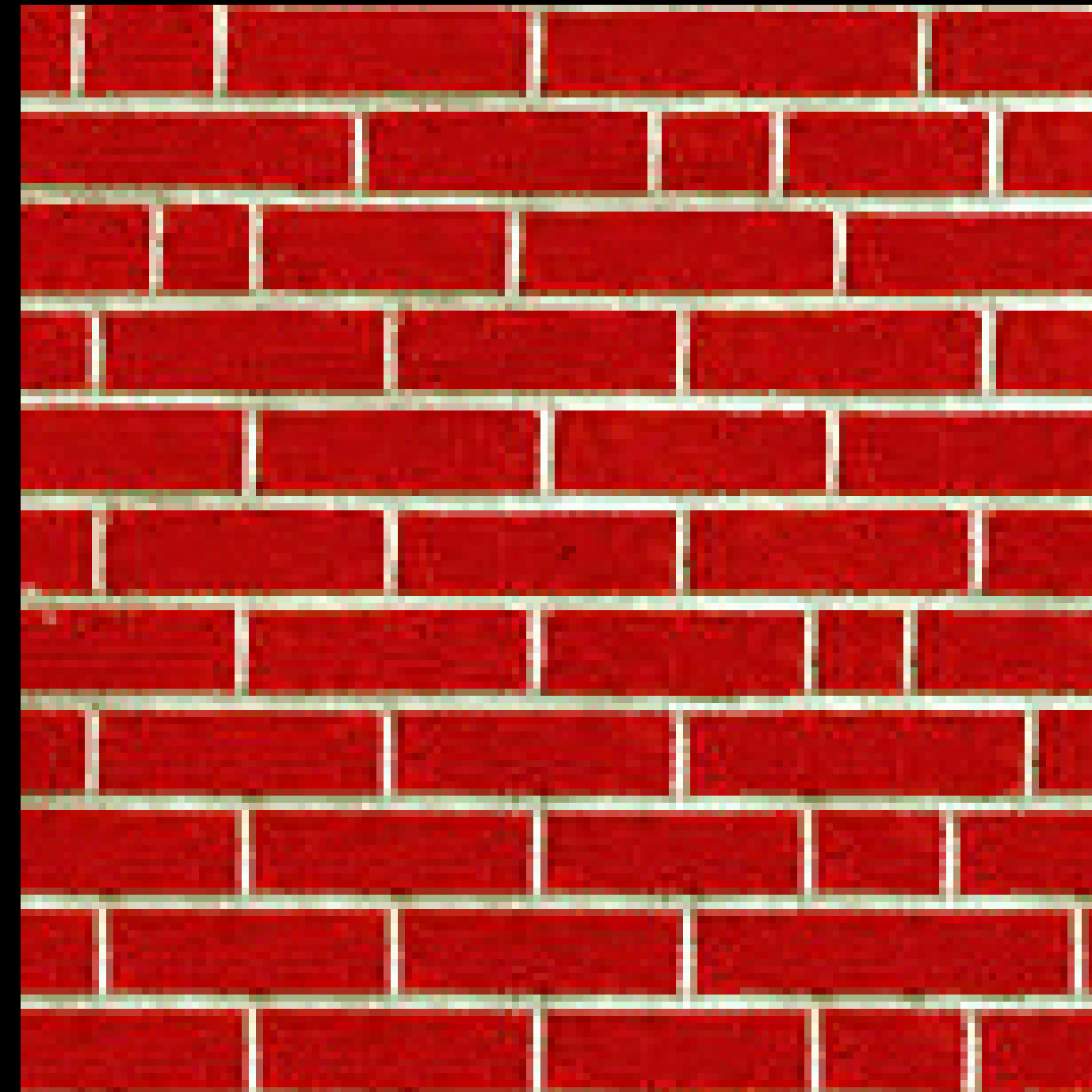
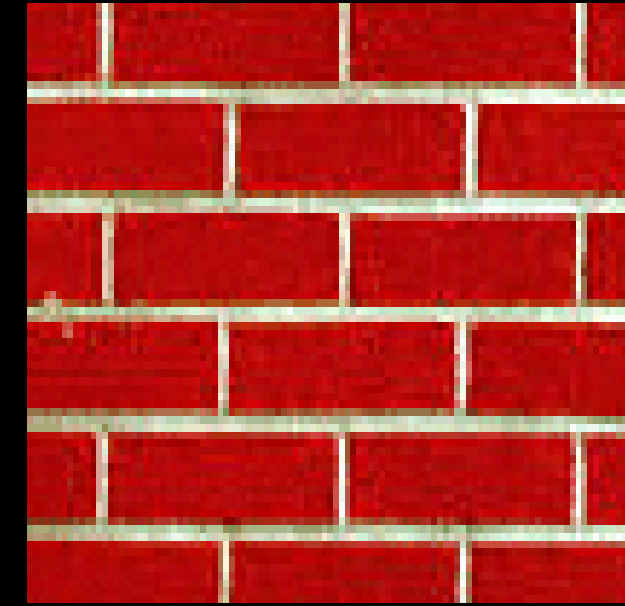


More Results

white bread

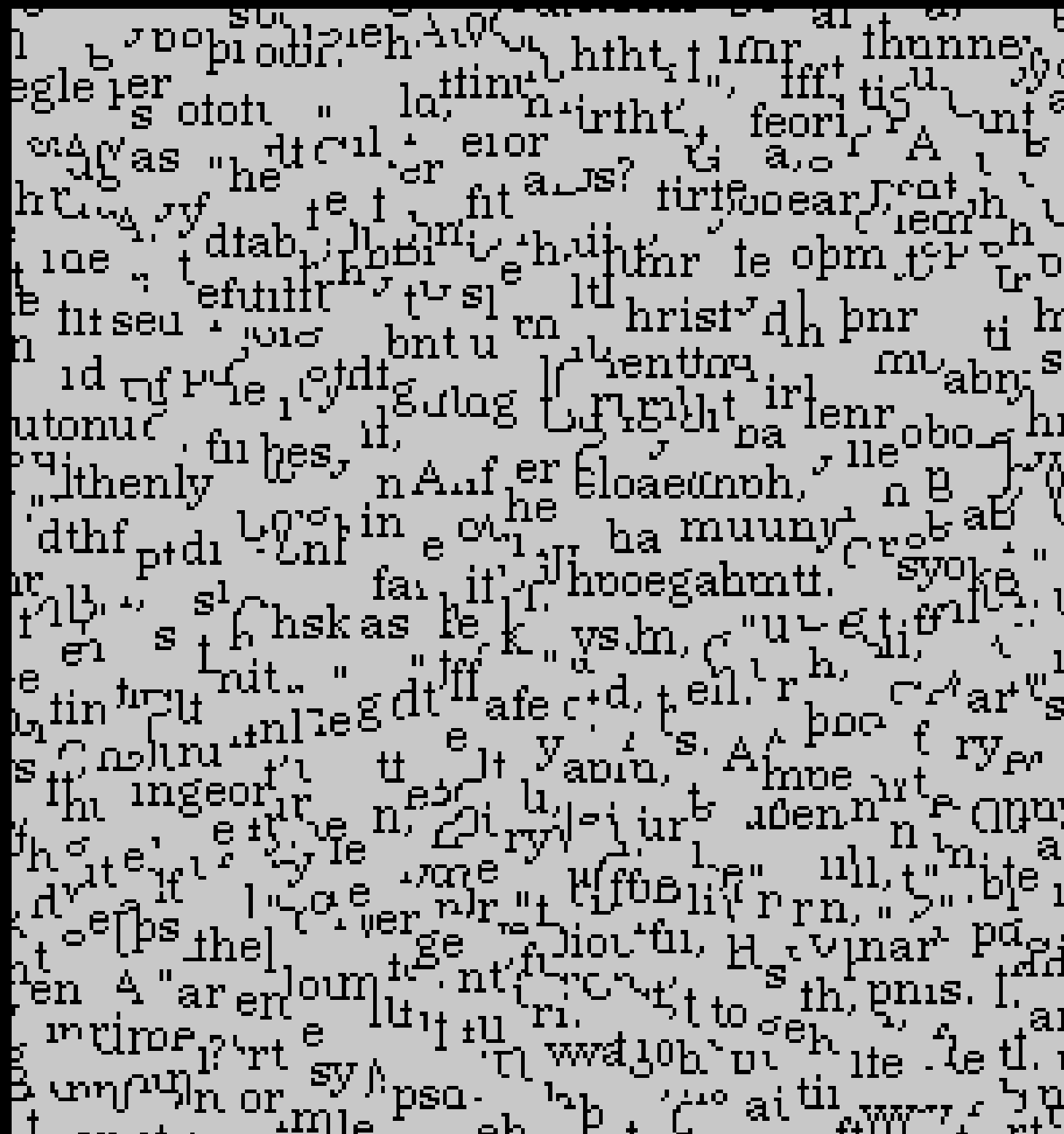


brick wall



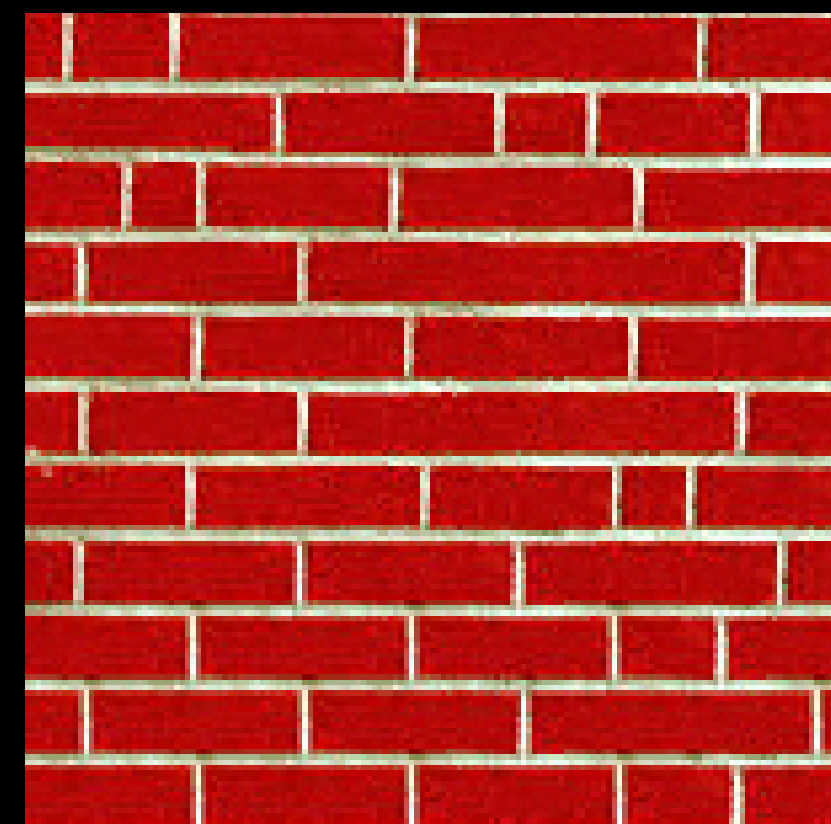
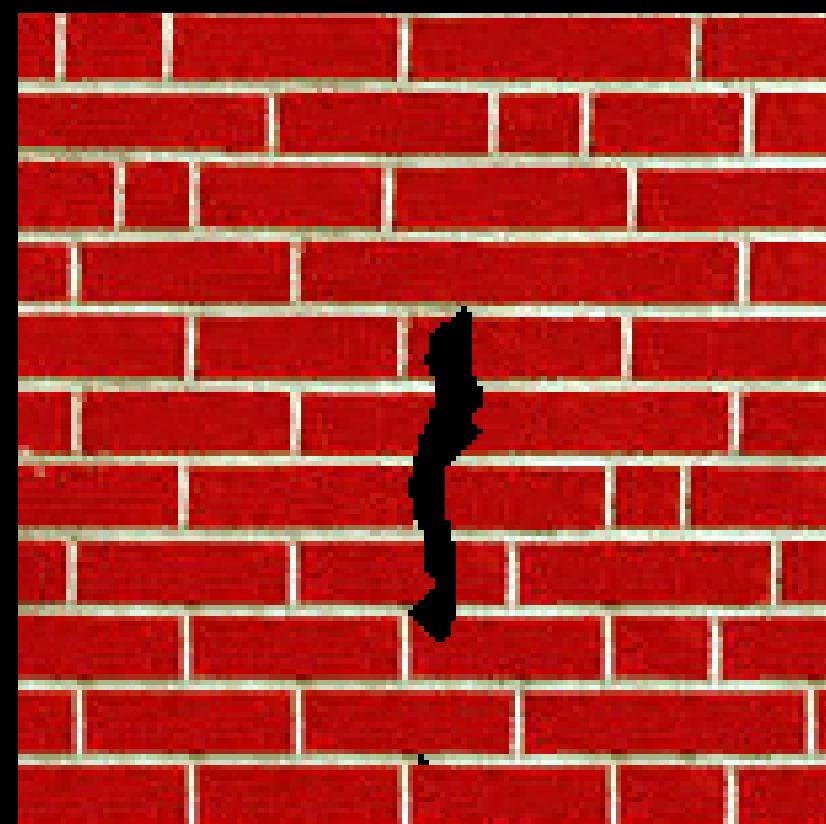
Homage to Shannon

...coming in the unsensational
...r Dick Gephardt was fai
...rful riff on the looming
...nly asked, "What's your
...tions?" A heartfelt sigh
...story about the emergen
...es against Clinton. "Boy
...g people about continuin
...hardt began, patiently obs
...s, that the legal system h
...g with this latest tanger



...athaim. them. "Whnephartfe lartifelintomimer
...fel ck Clirticout omaim thartfelins.f out s anetc
...the ry onst wartfe lck Gephtoomimeationl sigab
...Chicoufit Clinut Clil riff on. hat's yordn, parut tly
...ons yontonsteht waked, paim t sahe loo riff on
...nskoneploourtfeas leil A nst Clit, "Wleontongal s
...k Clirticouirtfepe.ong pme abegal fartfenstemem
...tiensteneltorydt telemephminsberdt was agemer
...ff ons artientont Cling peme as irtfe atih, "Boui s
...nal s fartfelt sig pedrtthdt ske abounutie aboutioo
...tfeone was you aboxonhardt thatins fain, ped, '
...ains. them, pabout wasy arfnt coutly d, l n A h
...ole emthrdngbooreme agas fa bontinsyst Clinut
...ory about continst Clipeopinst Cloke agatiff out C
...stome zinemen tly ardt beoraboul n, thenly as t C
...cons fairmeme Diontont wat coutlyohgans as fan
...ien, phrtfaul, "Wbaut cout congagal comiringa
...mifmst Clily abon al cocountha.emungairt tf ou
...The loocrystal loontieph. intly on, theoplegatick
...oul tatiezontly atie Diontiomt wal s f tbegea ener
...nthahgat's enenhhbas fan. "intchthorw abons w

Hole Filling



Extrapolation

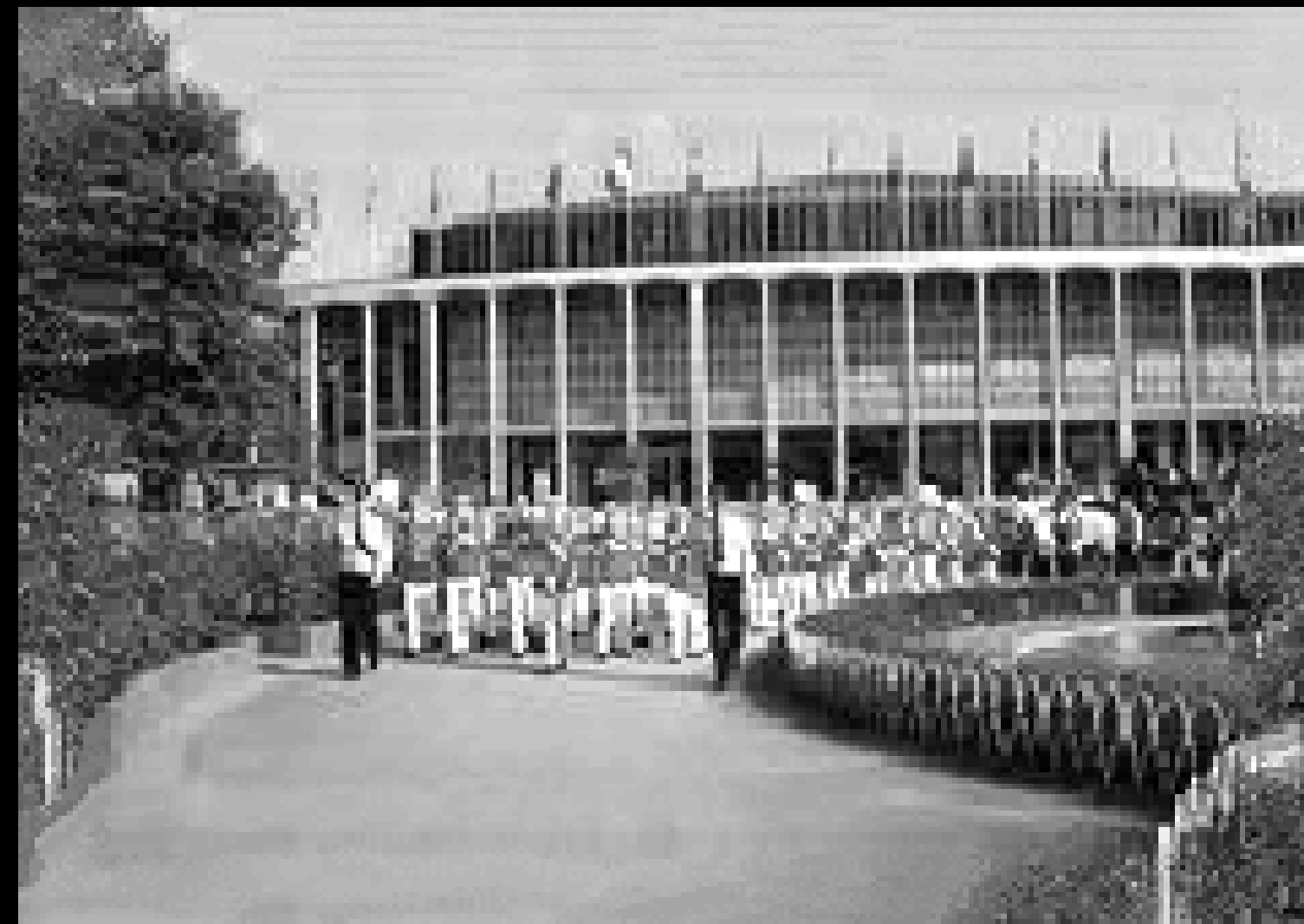
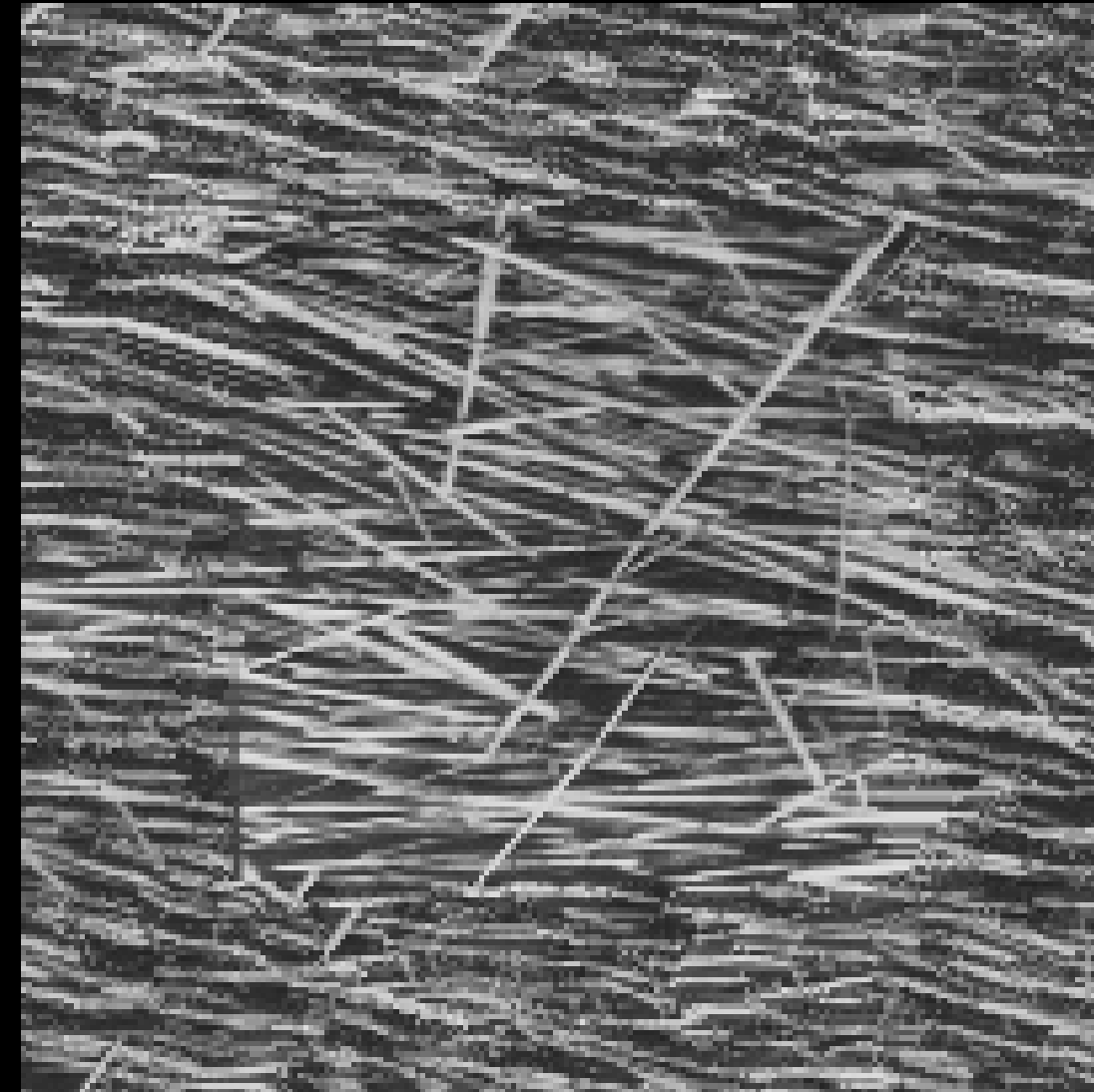
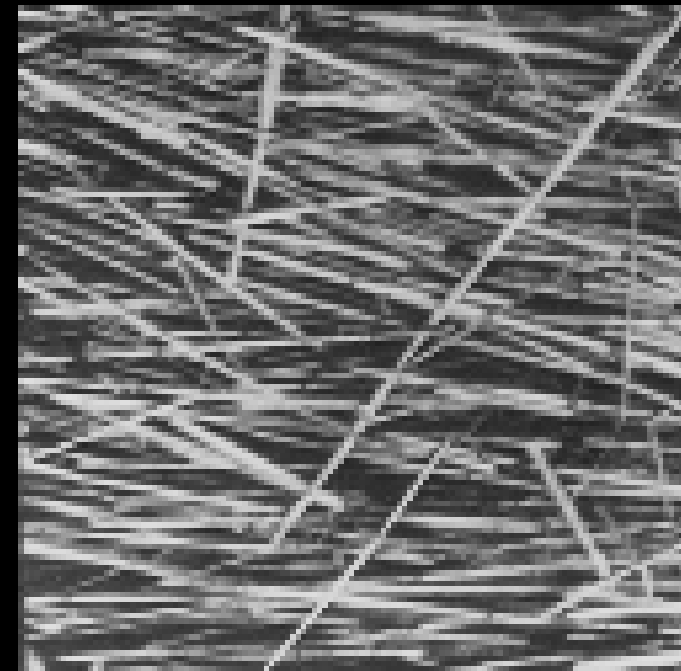


Image Analogies

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

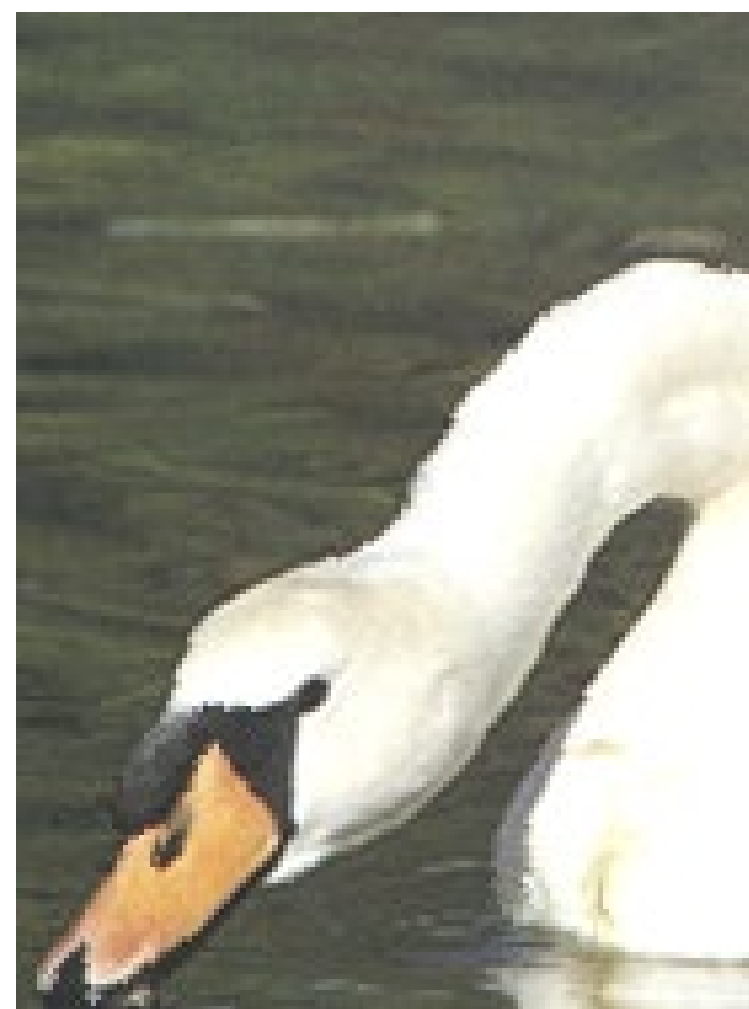


B'



Image Analogies

Goal: Process an image by example



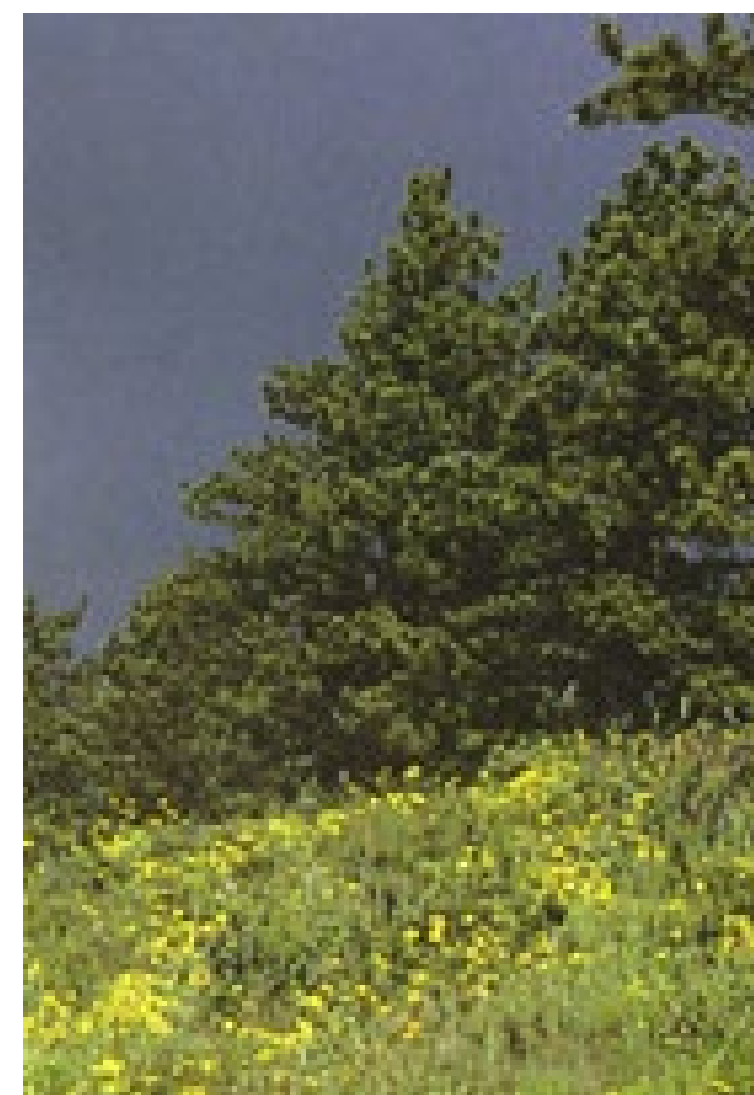
A

:



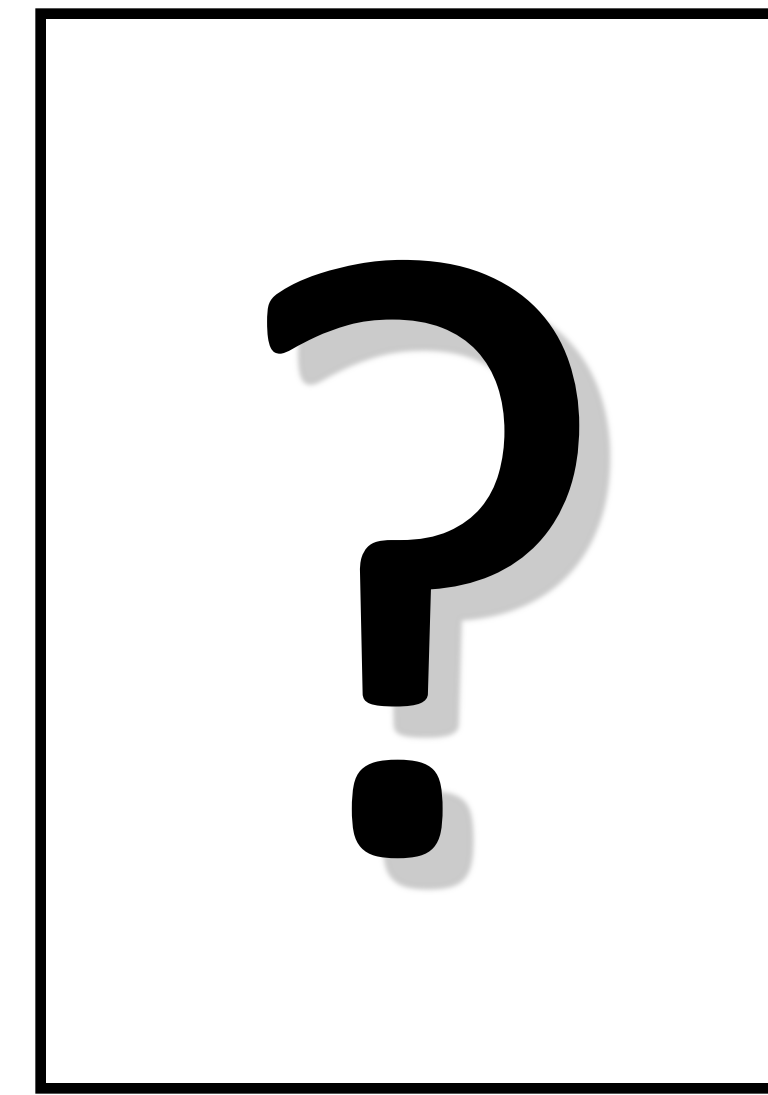
A'

::



B

:



B'

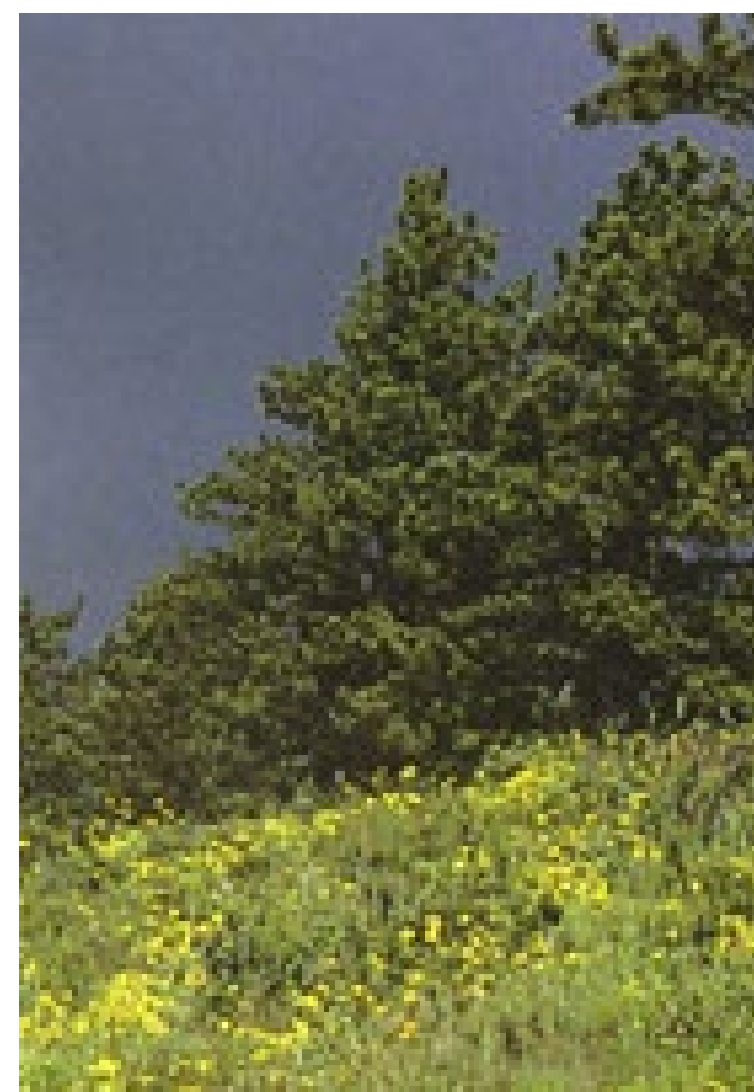
Non-parametric sampling



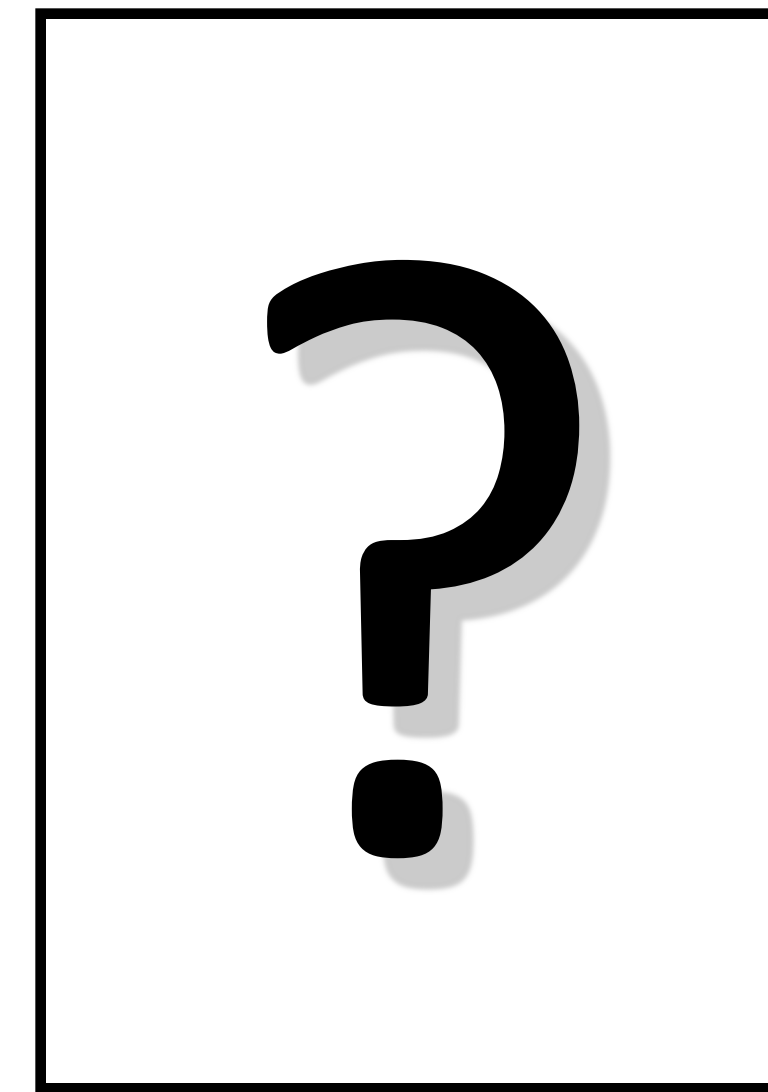
A



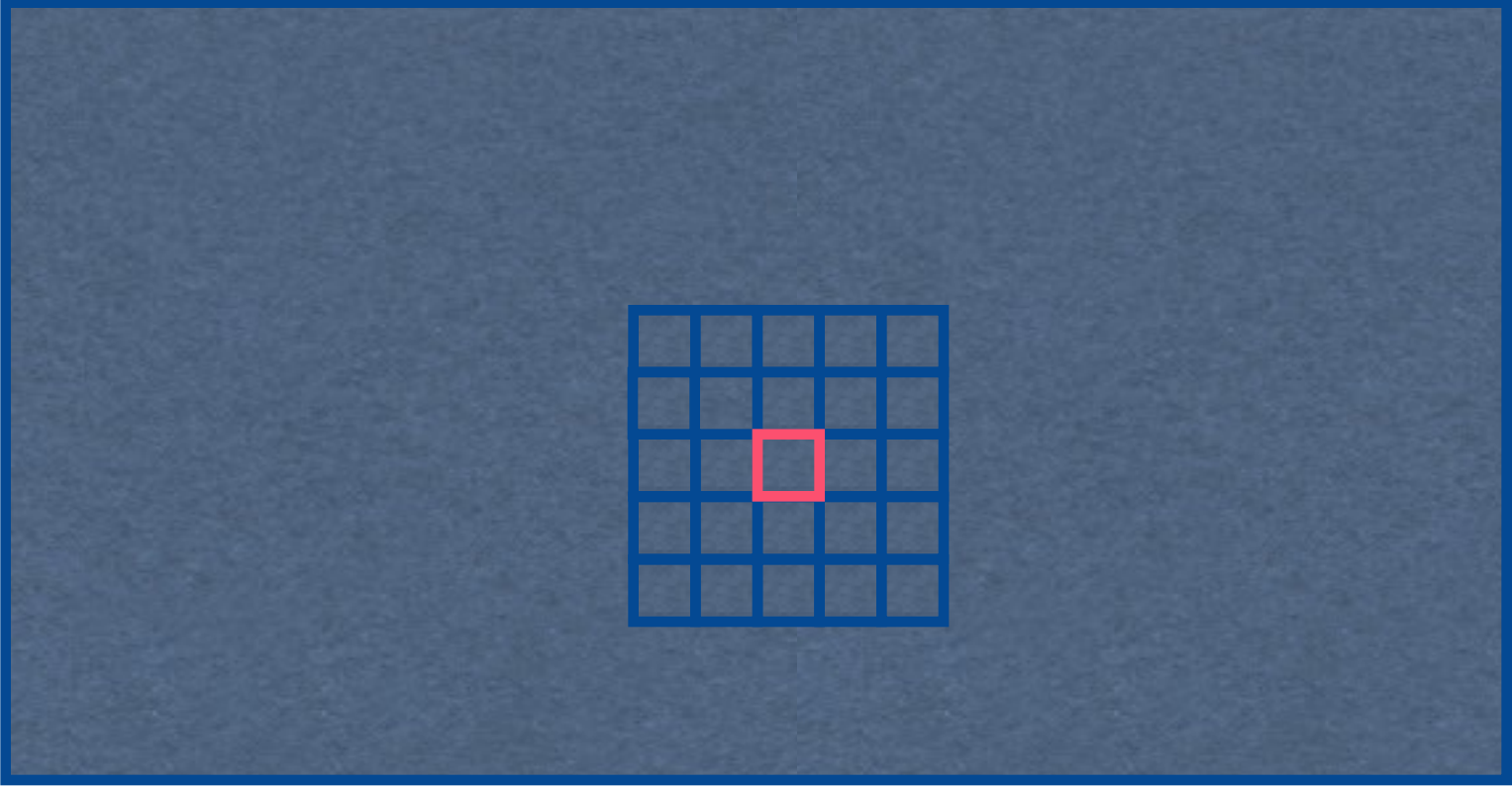
A'



B

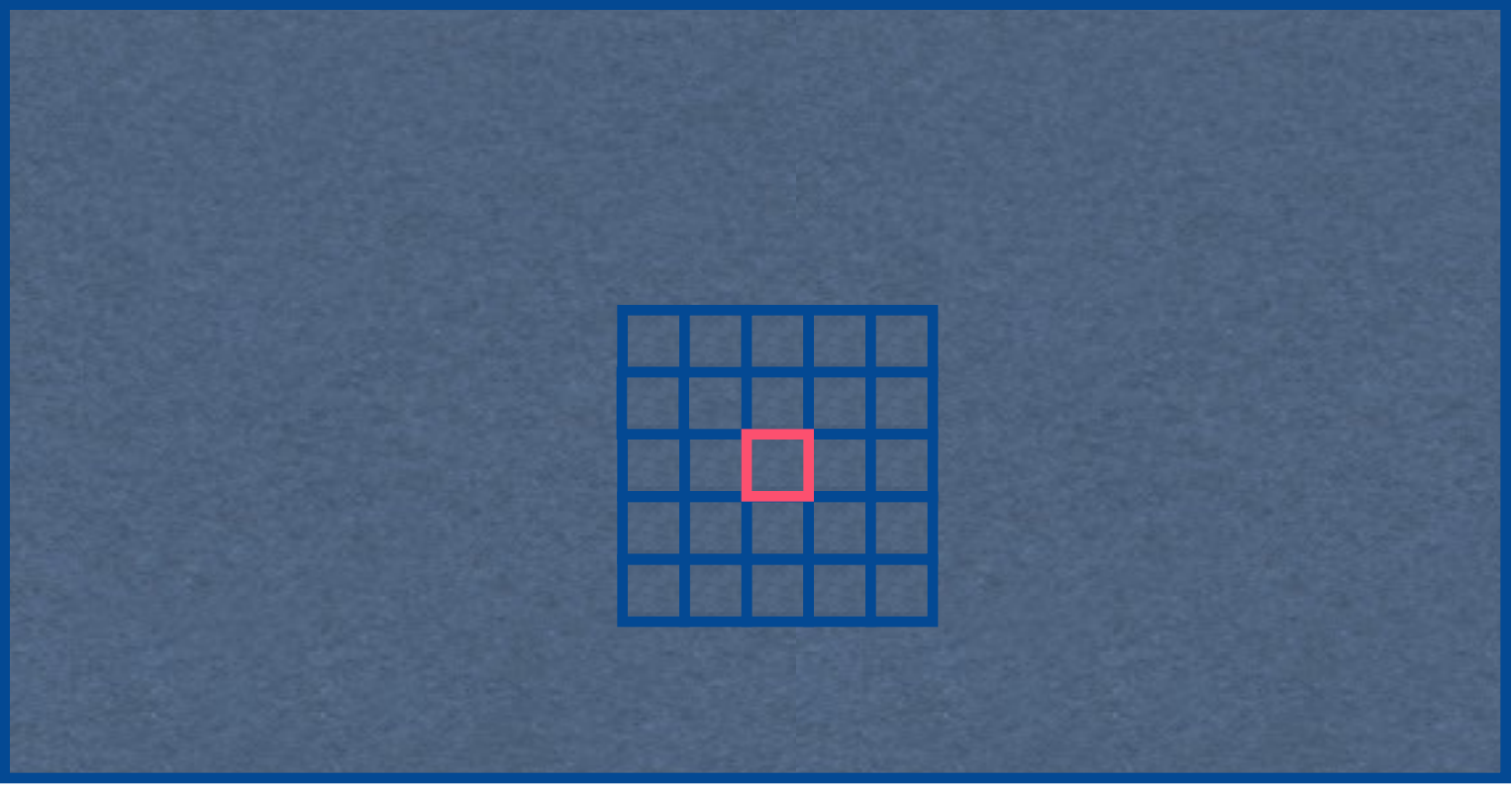


B'



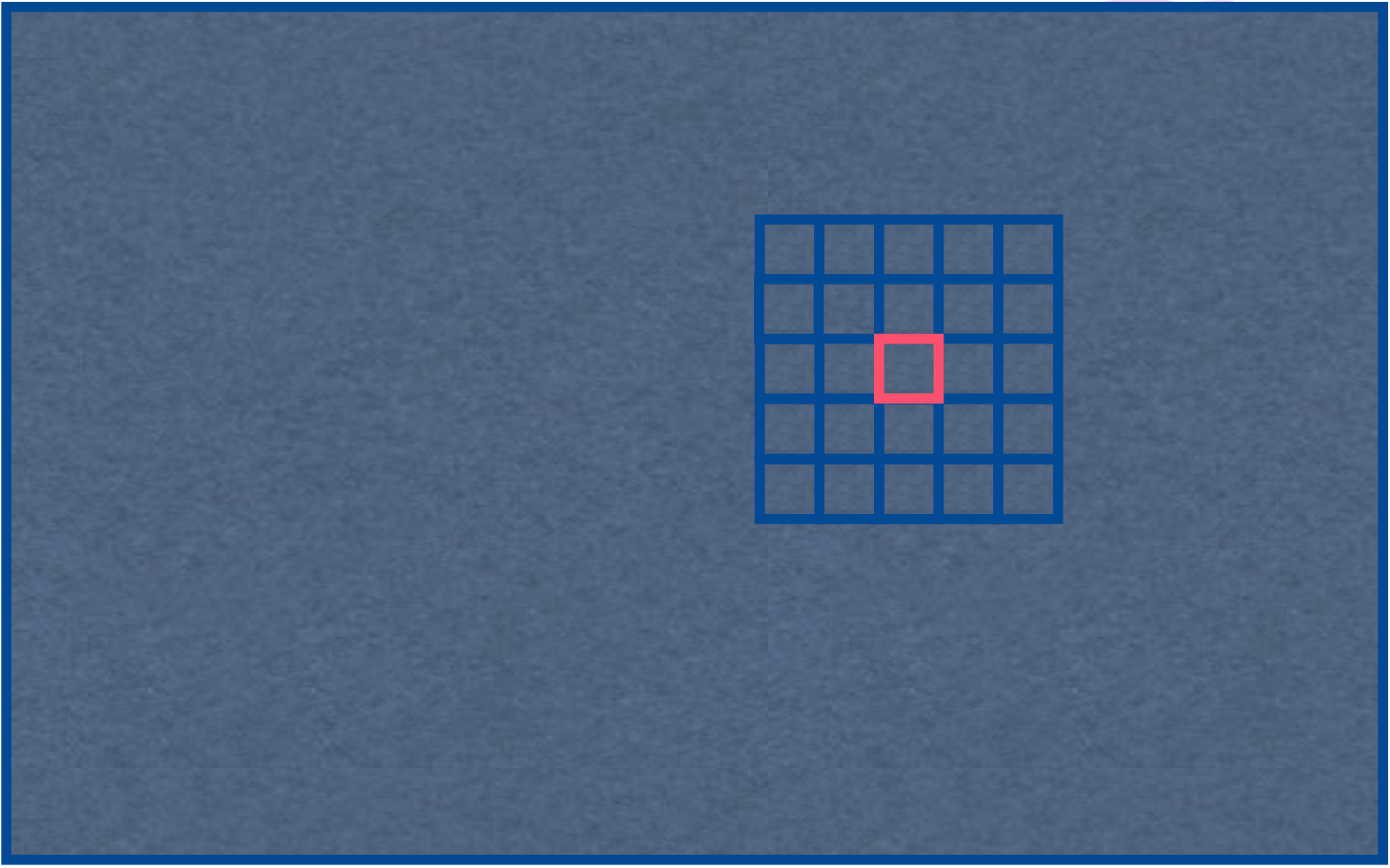
A

⋮



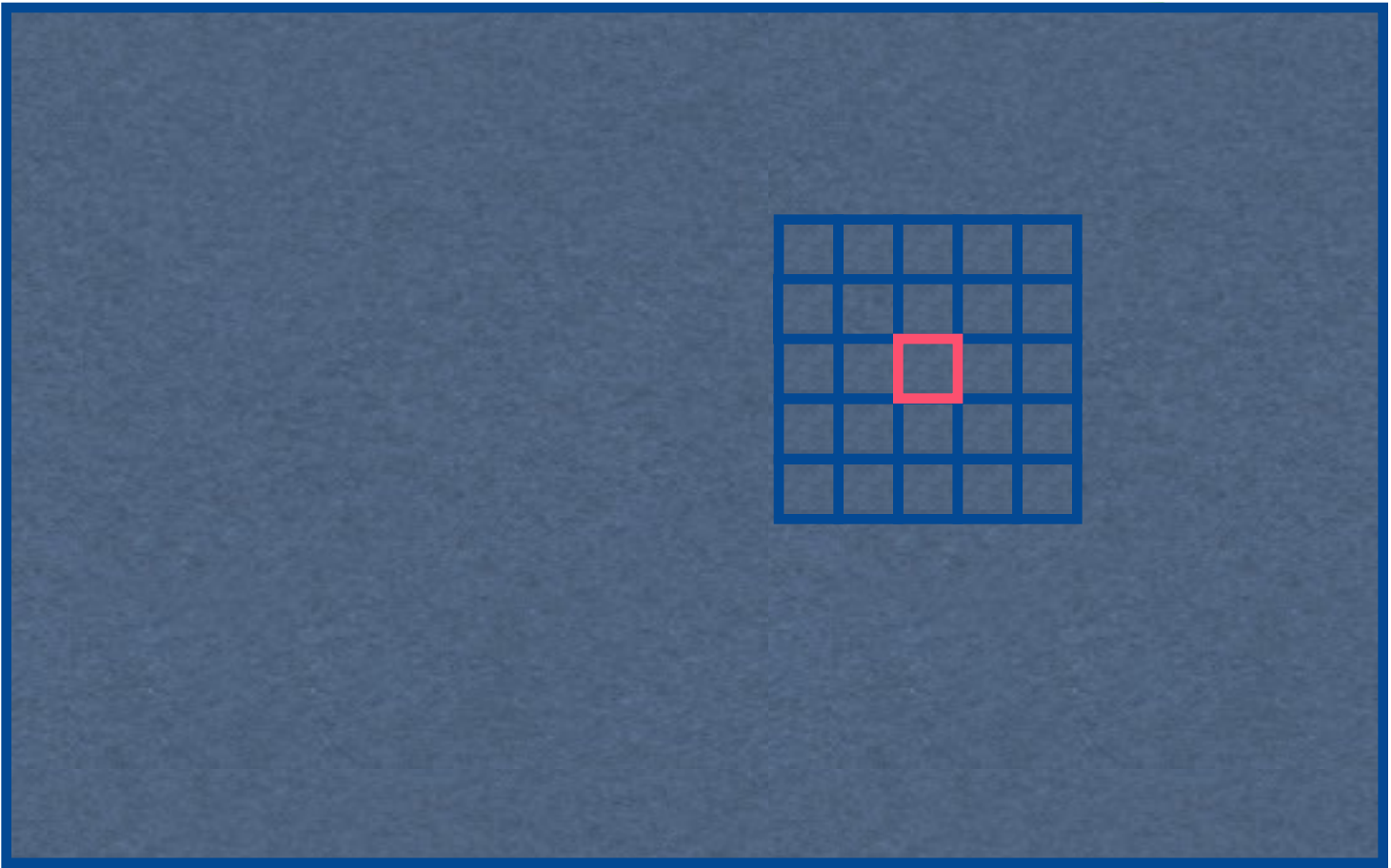
A'

⋮

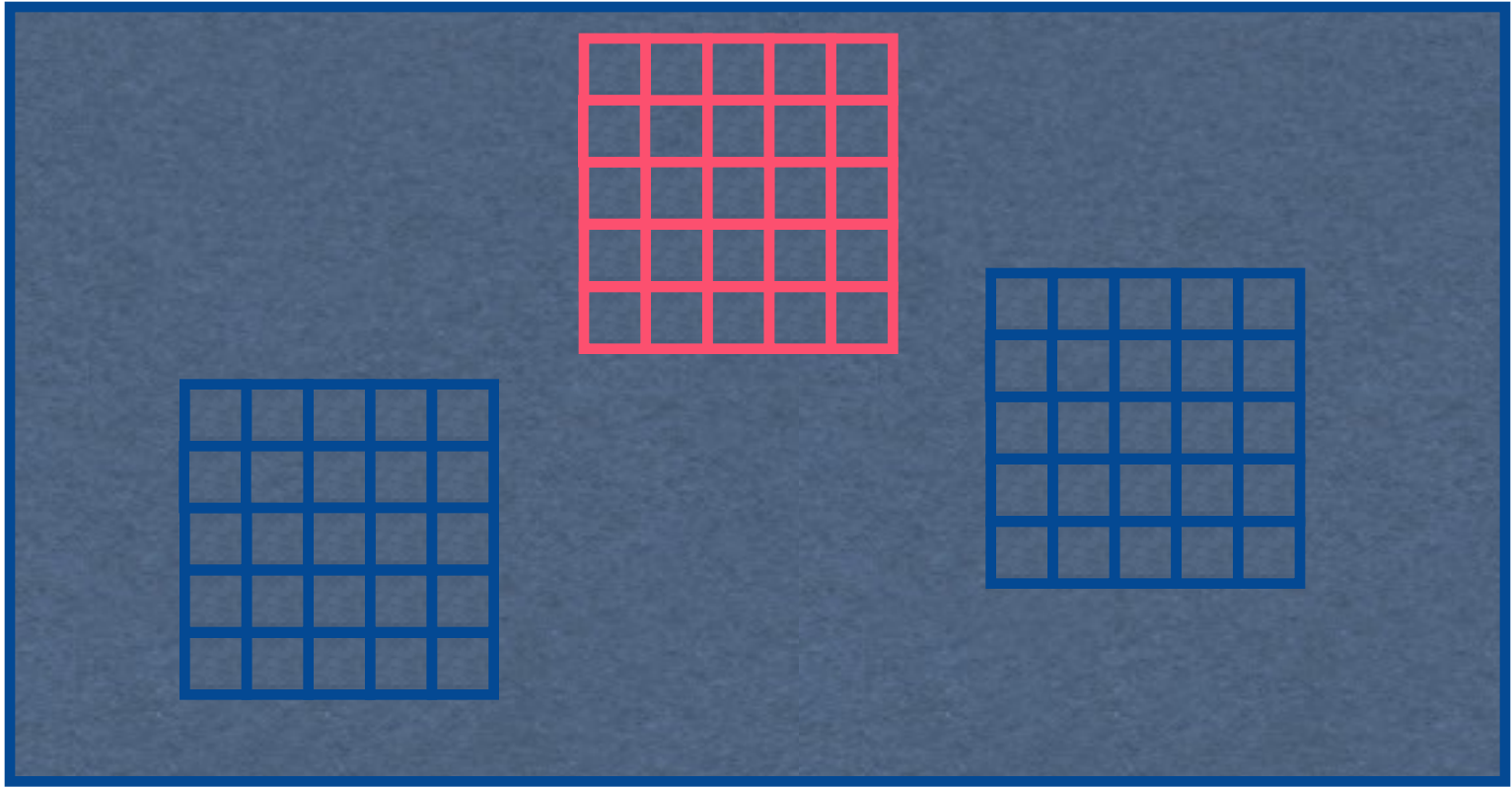


B

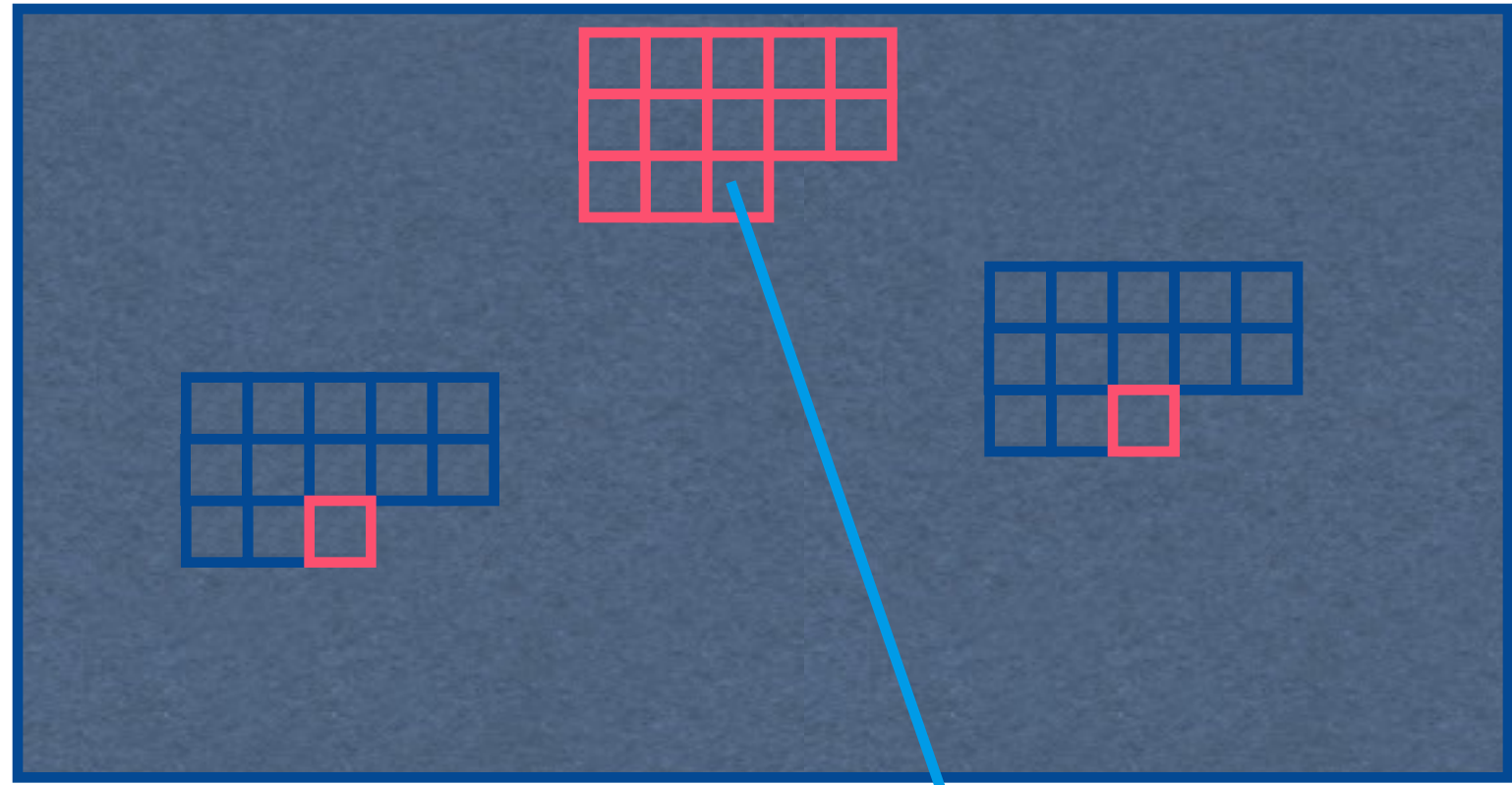
⋮



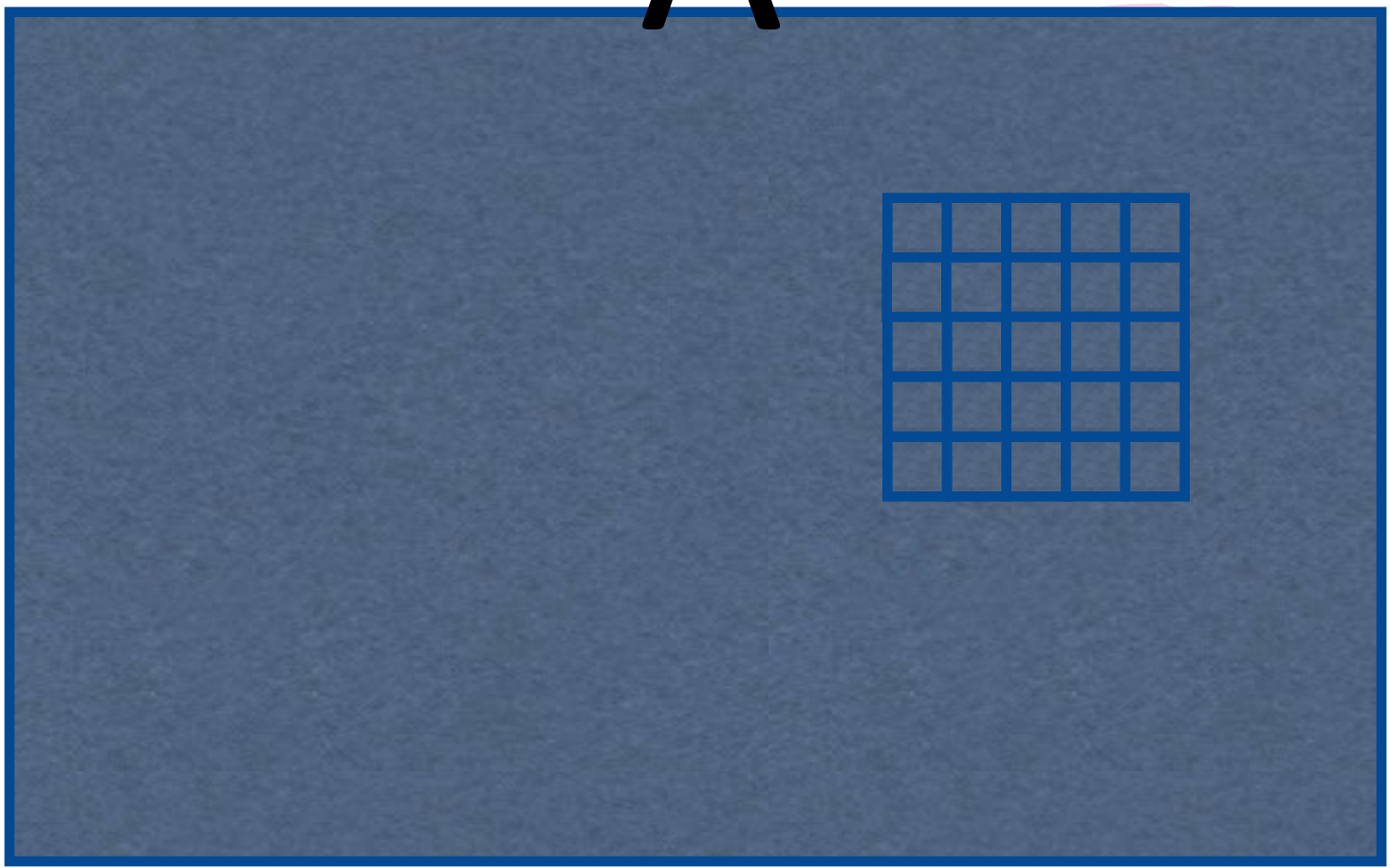
B'



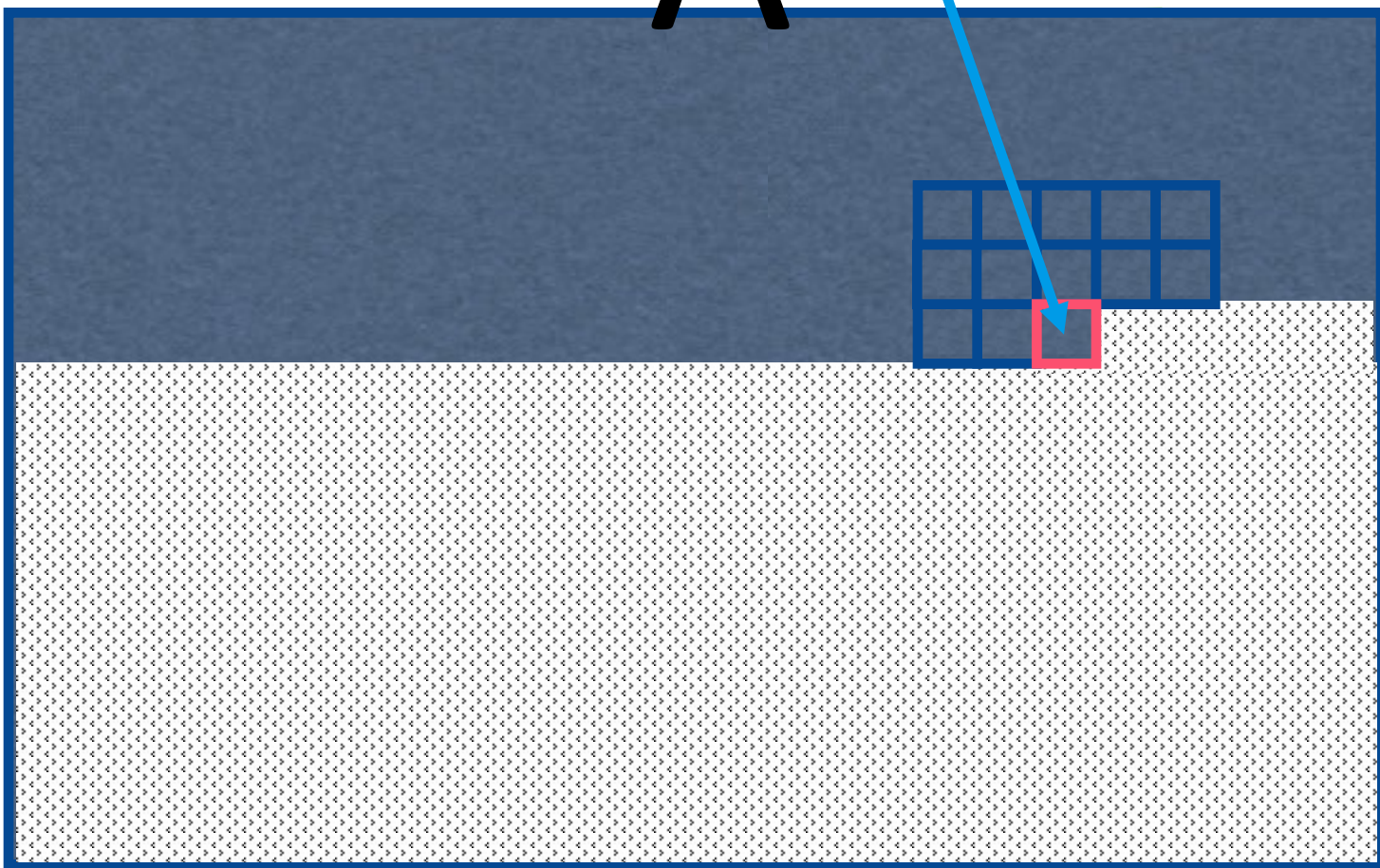
A



A'



B



B'

•
•

• •
• •

•
•

Blur Filter



Unfiltered source (A)



Filtered source (A')



Unfiltered target (B)

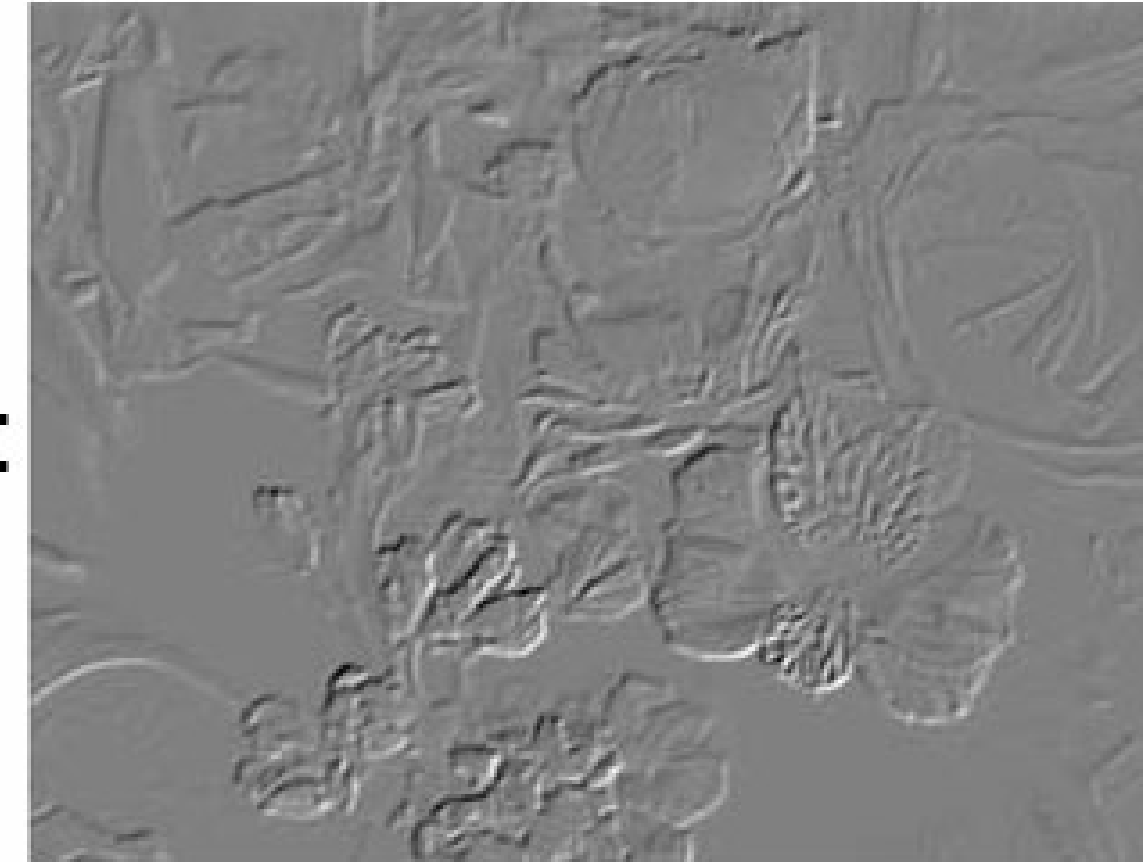


Filtered target (B')

Edge Filter



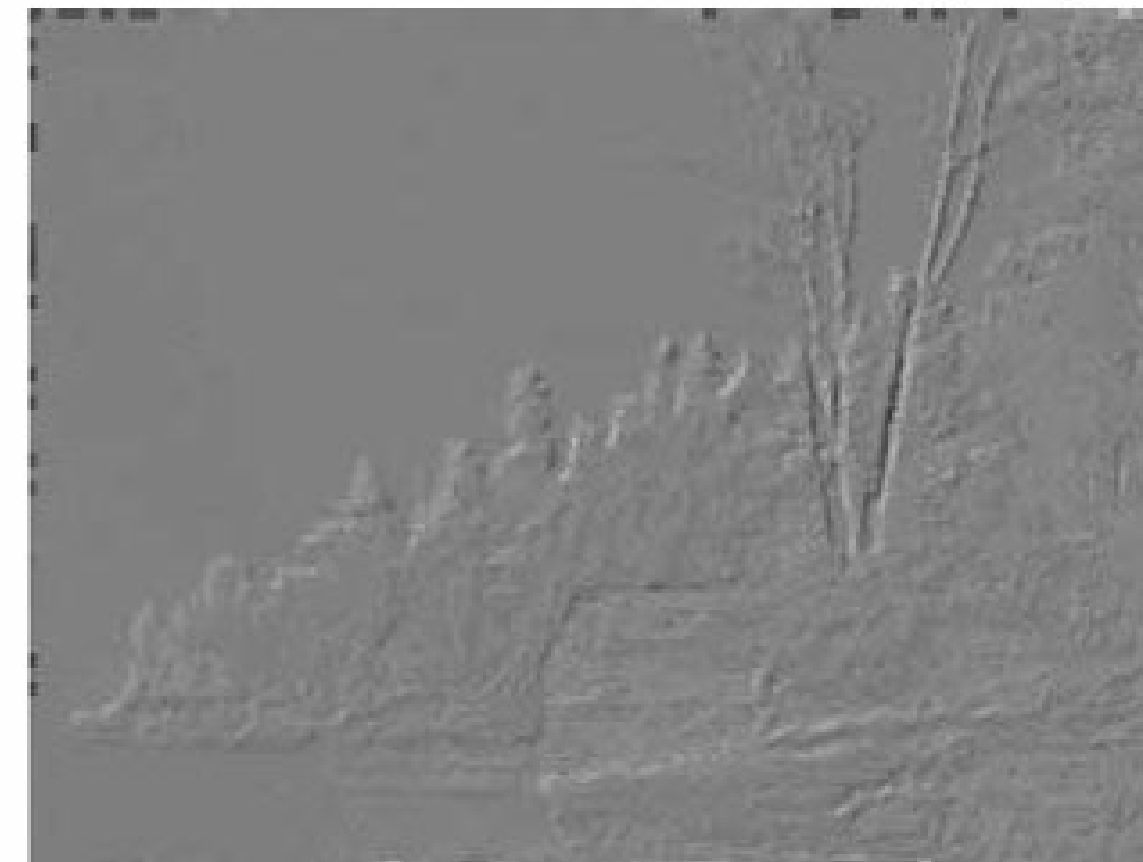
Unfiltered source (A)



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')

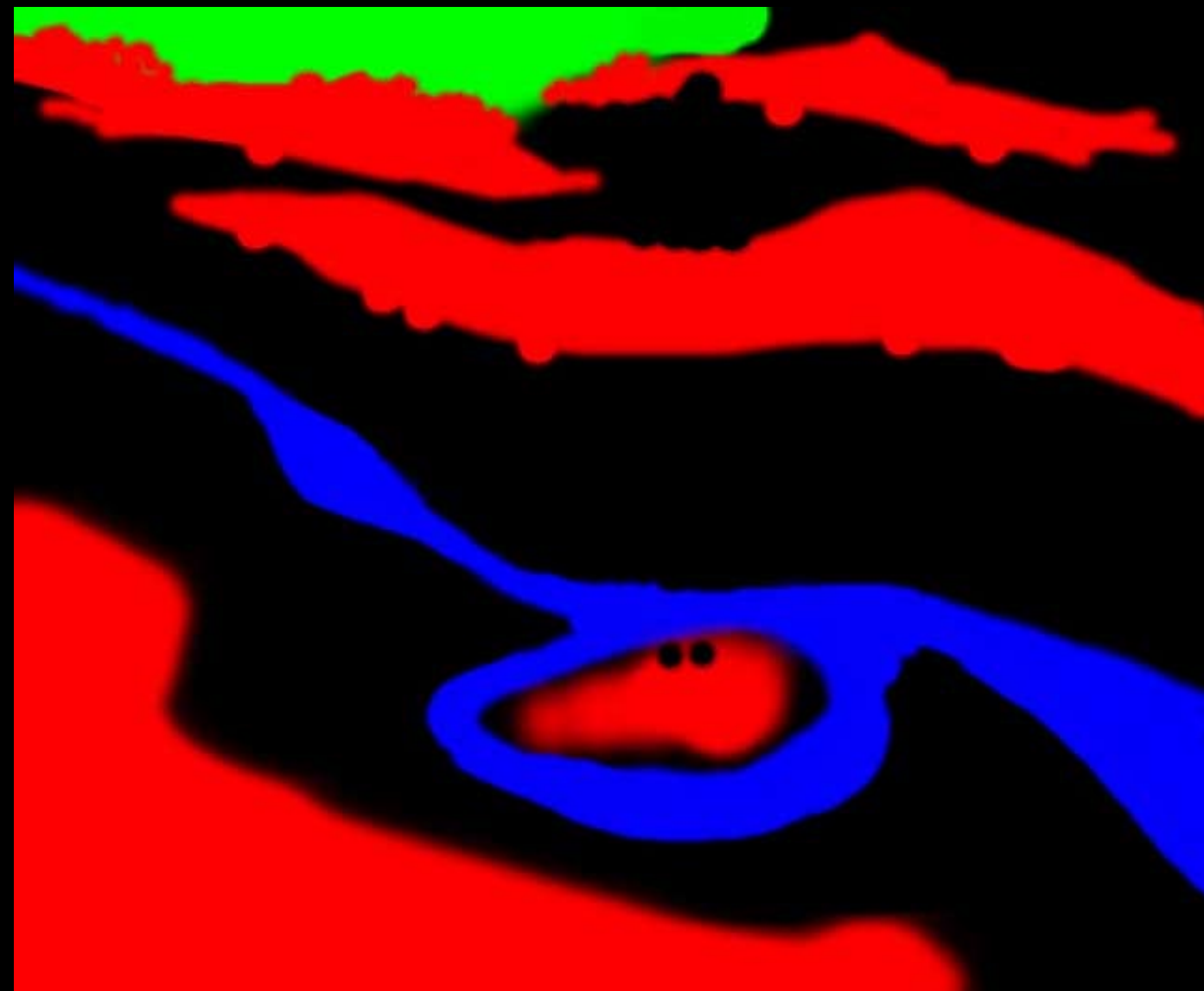
Texture-by-numbers



A



A'



B



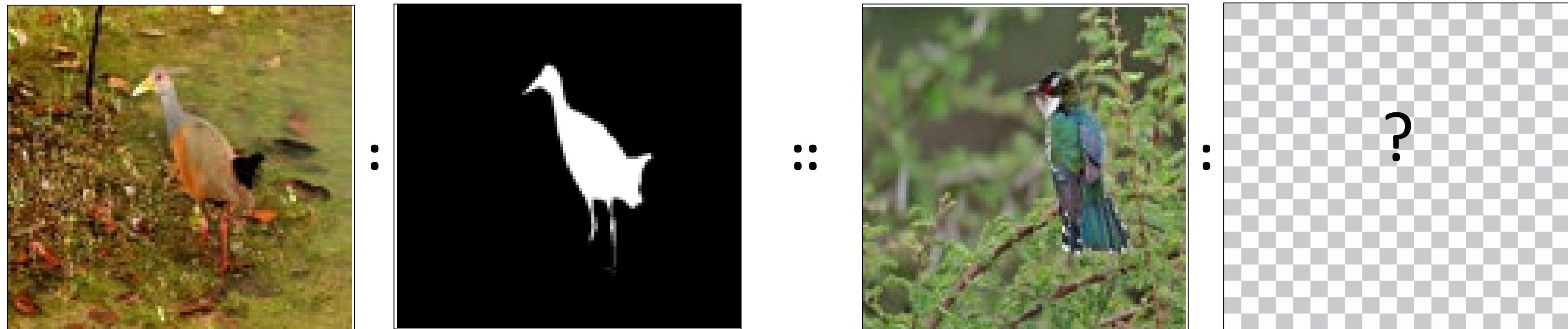
B'

Visual Prompting via Image Inpainting

Amir Bar*, Yossi Gandelsman*,

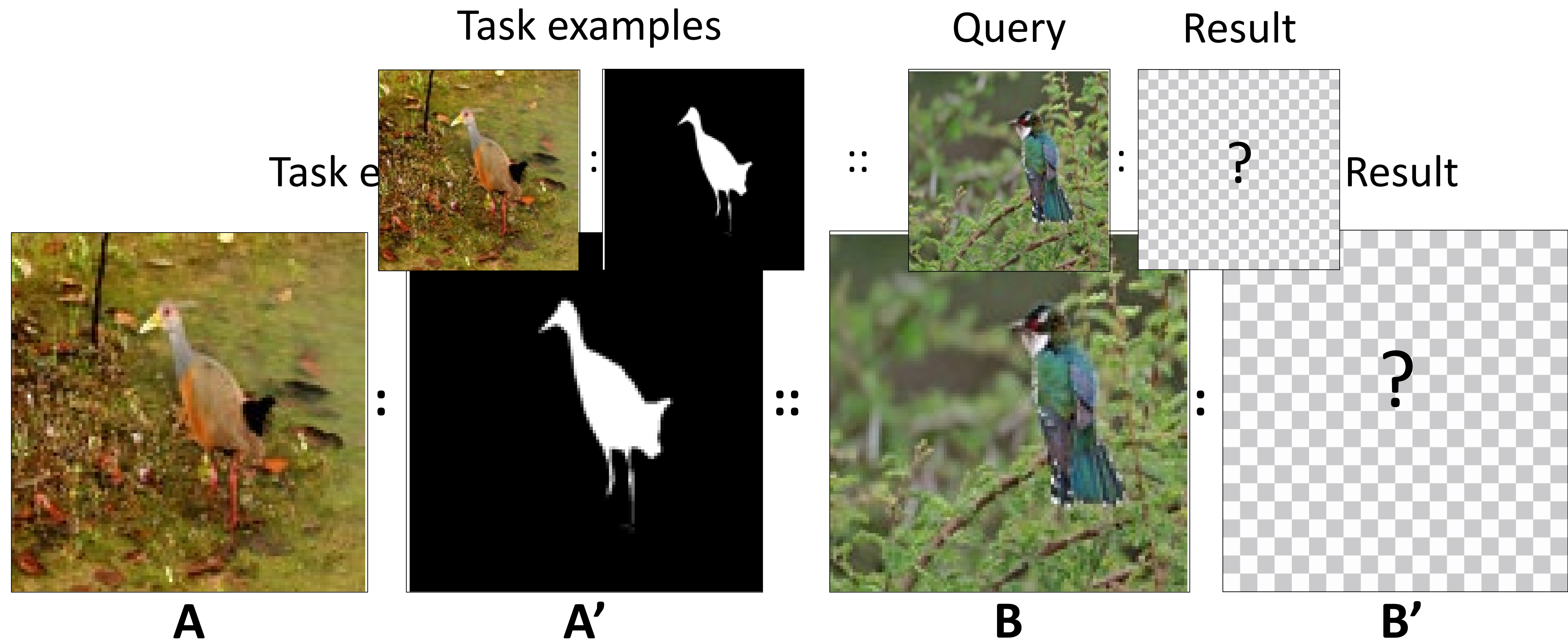
Trevor Darrell, Amir Globerson, Alexei A Efros

NeurIPS 2022



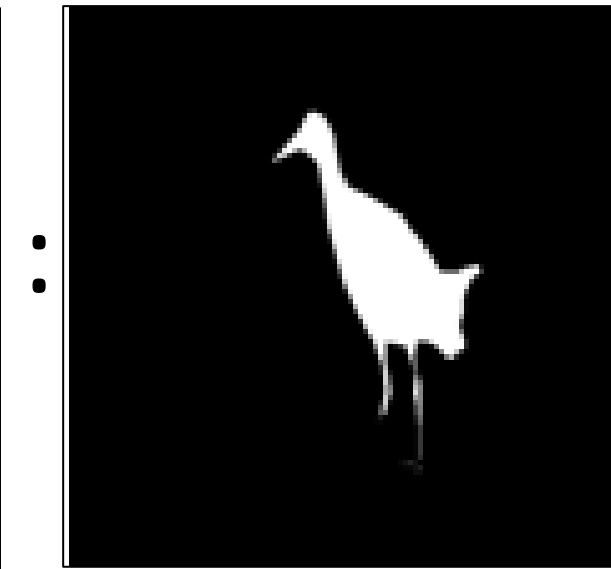
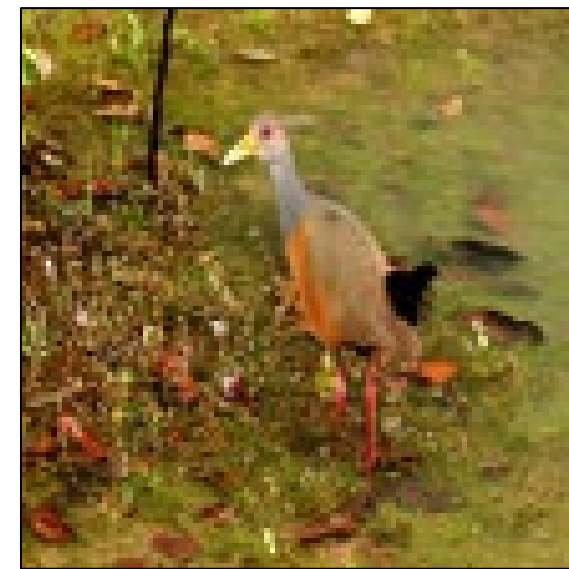
* Equal contribution

Visual Prompting



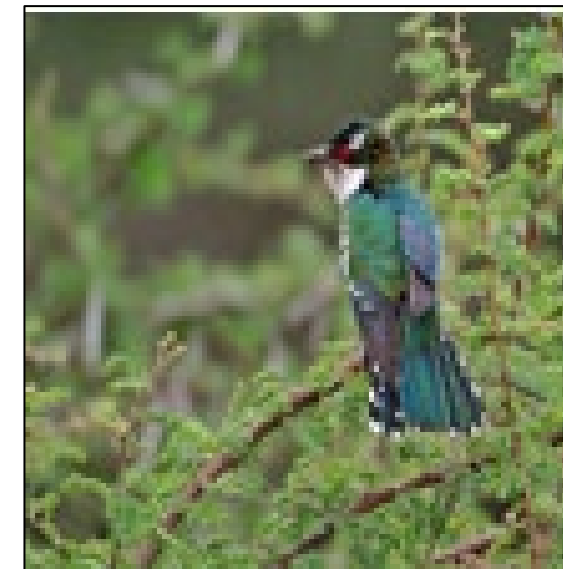
Visual Prompting

Task examples

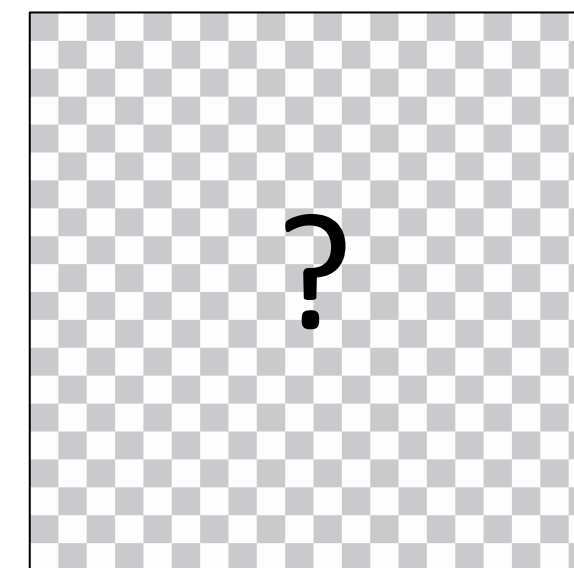
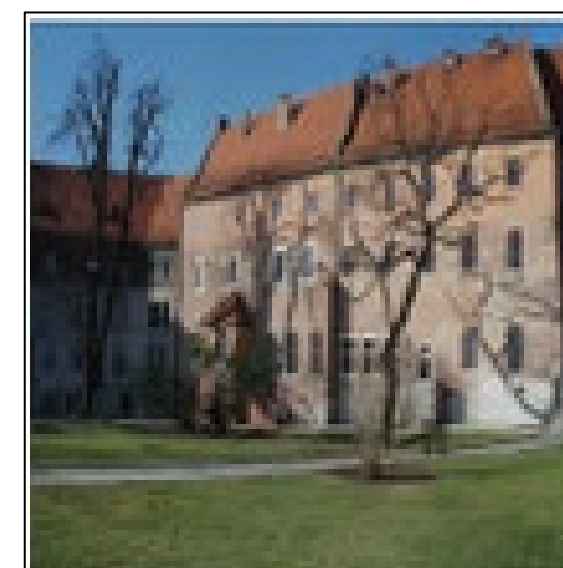
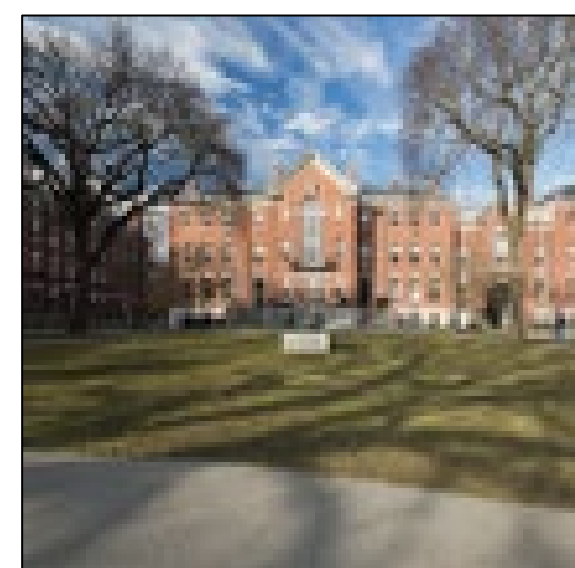
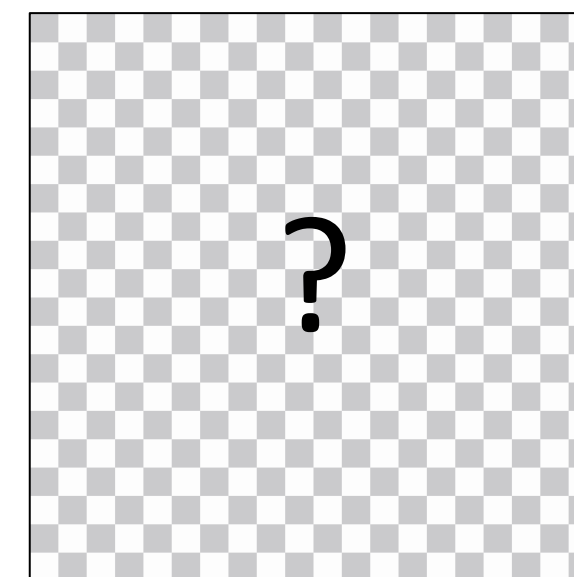
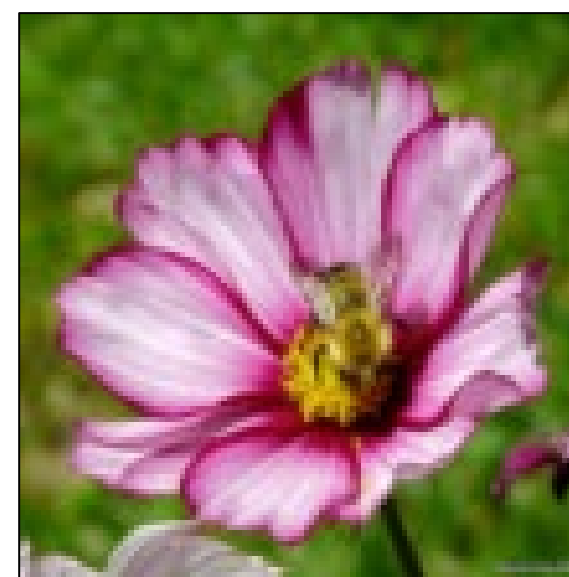
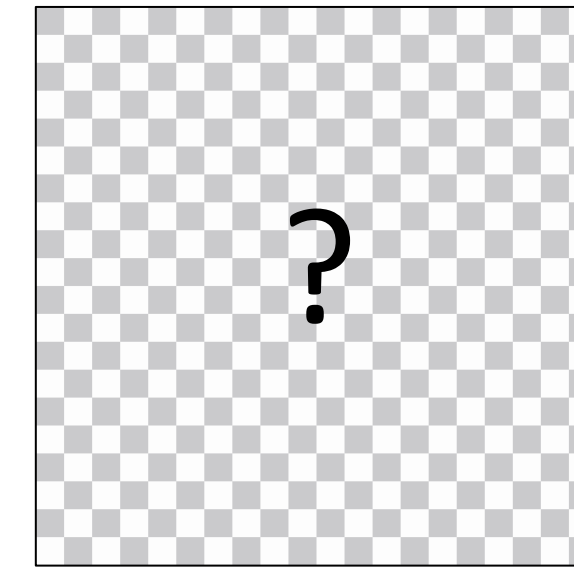


::

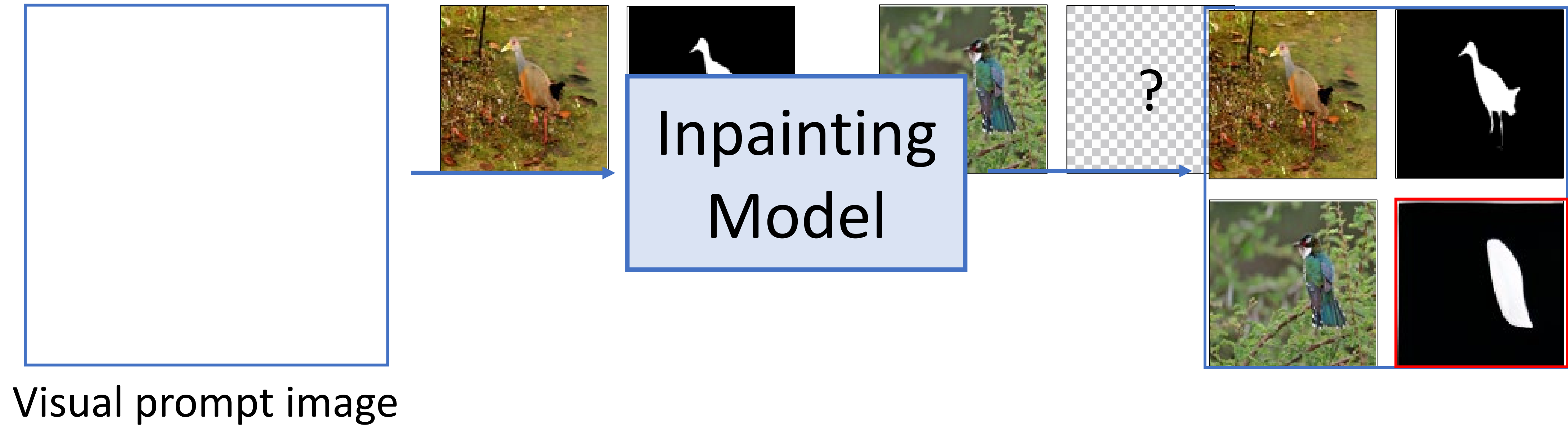
Query



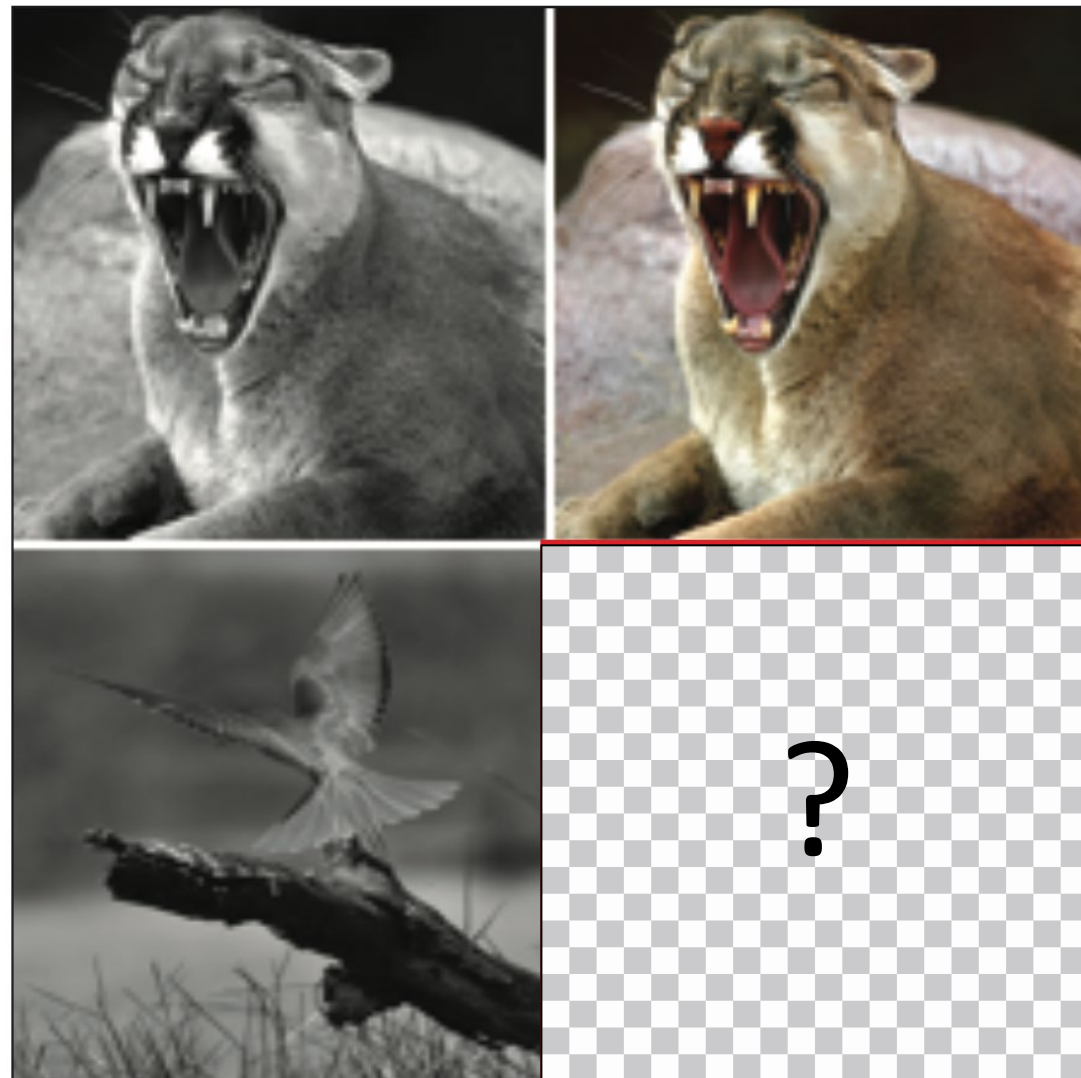
Result



Inpainting models to the rescue!



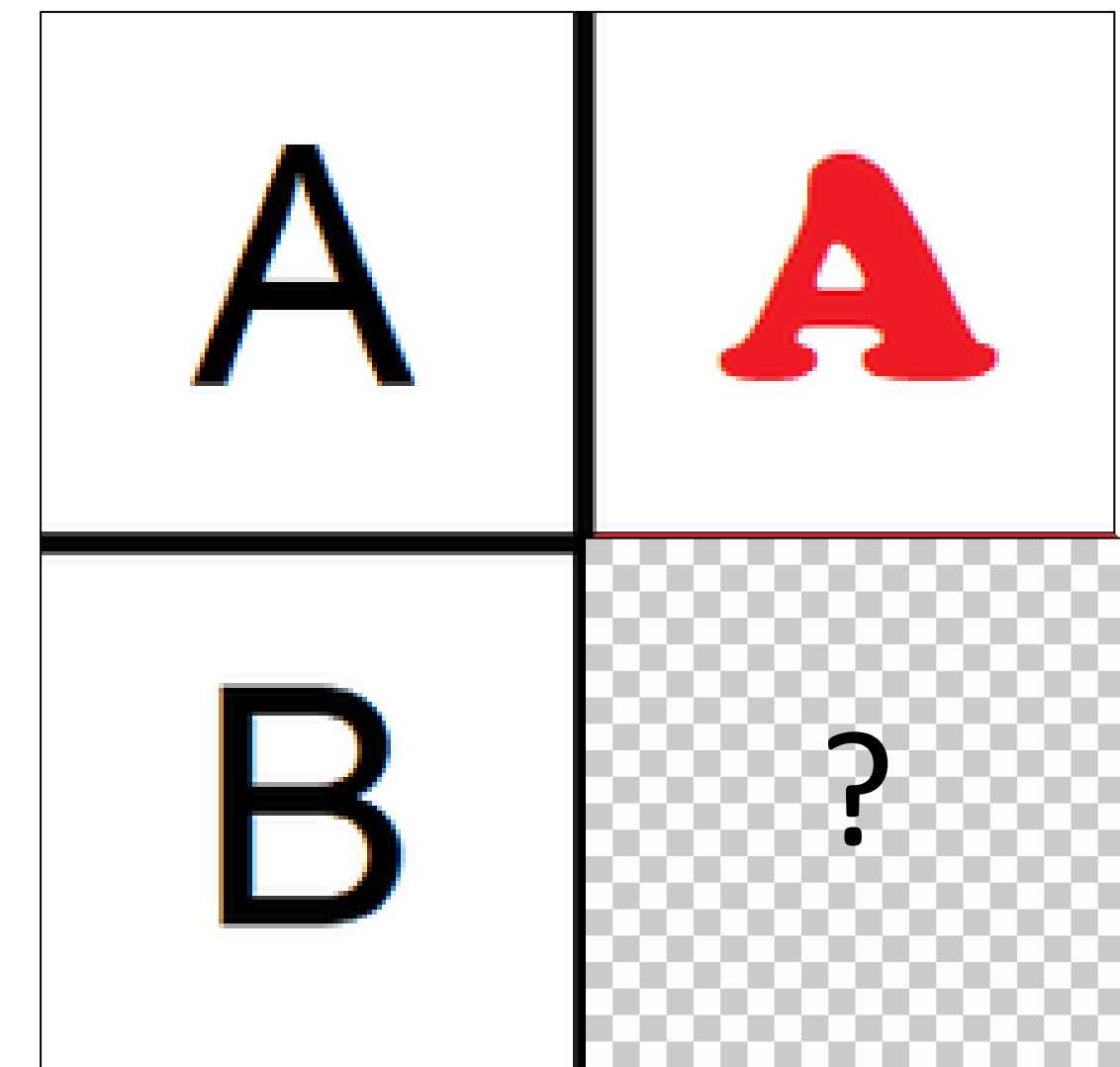
Wide range of tasks



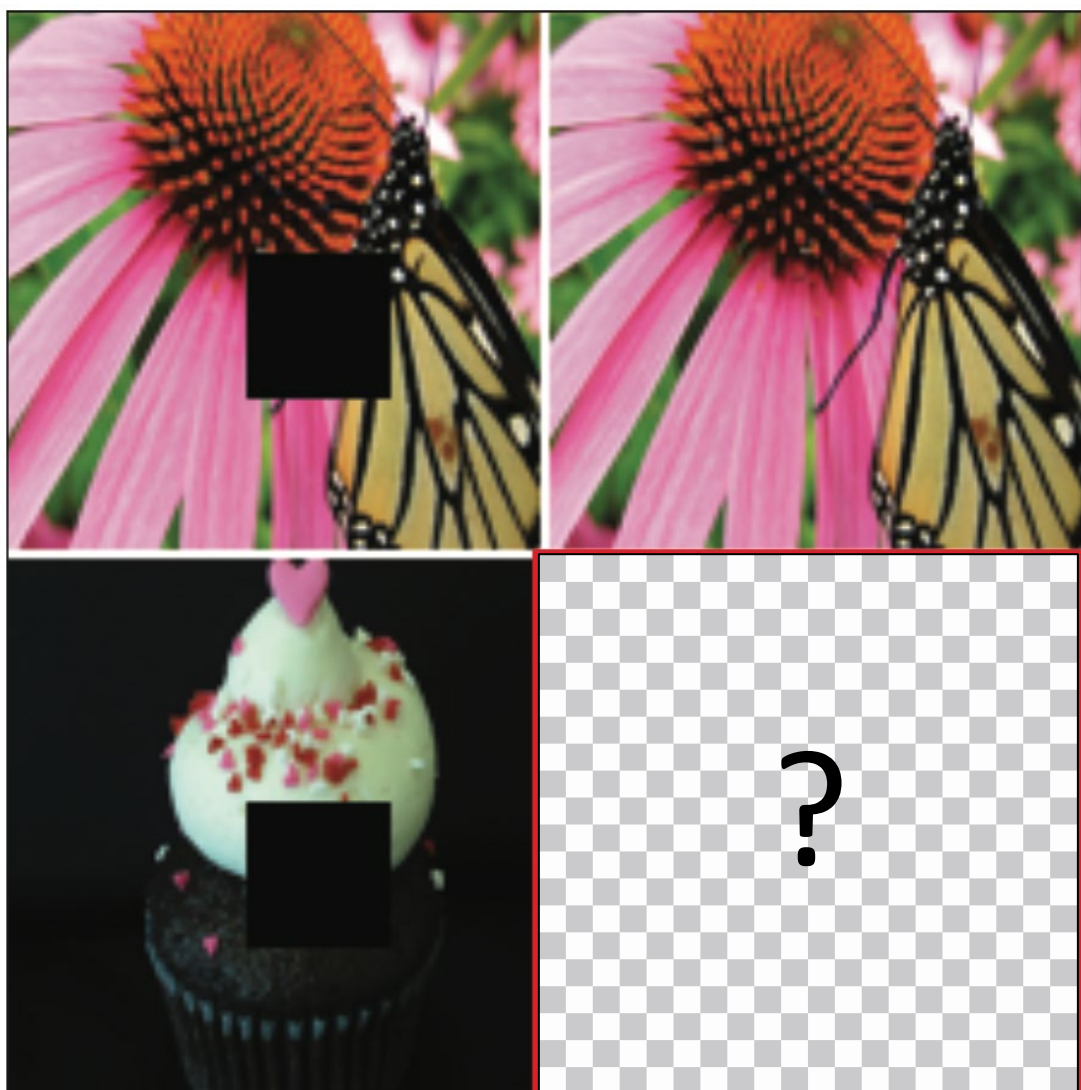
Colorization



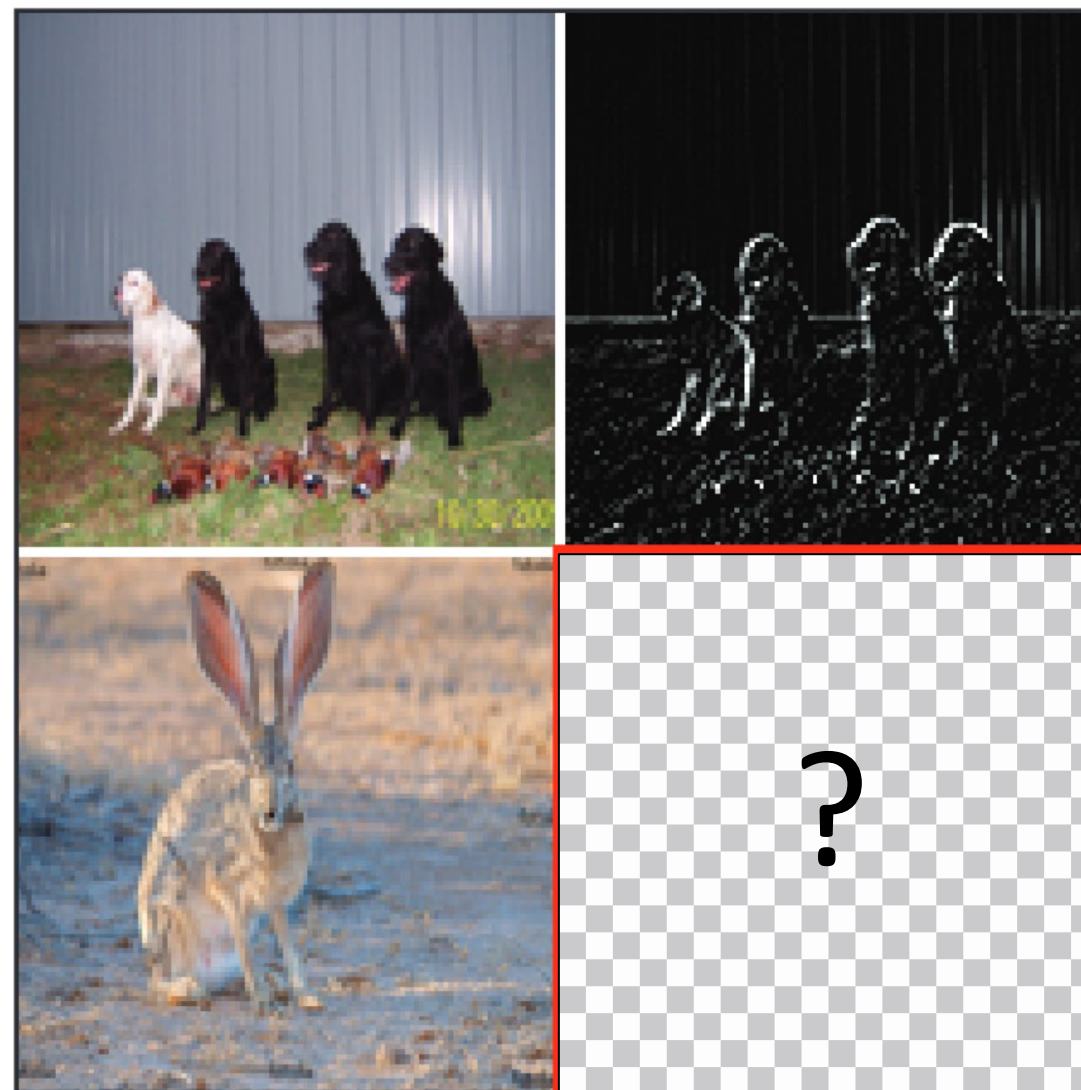
Segmentation



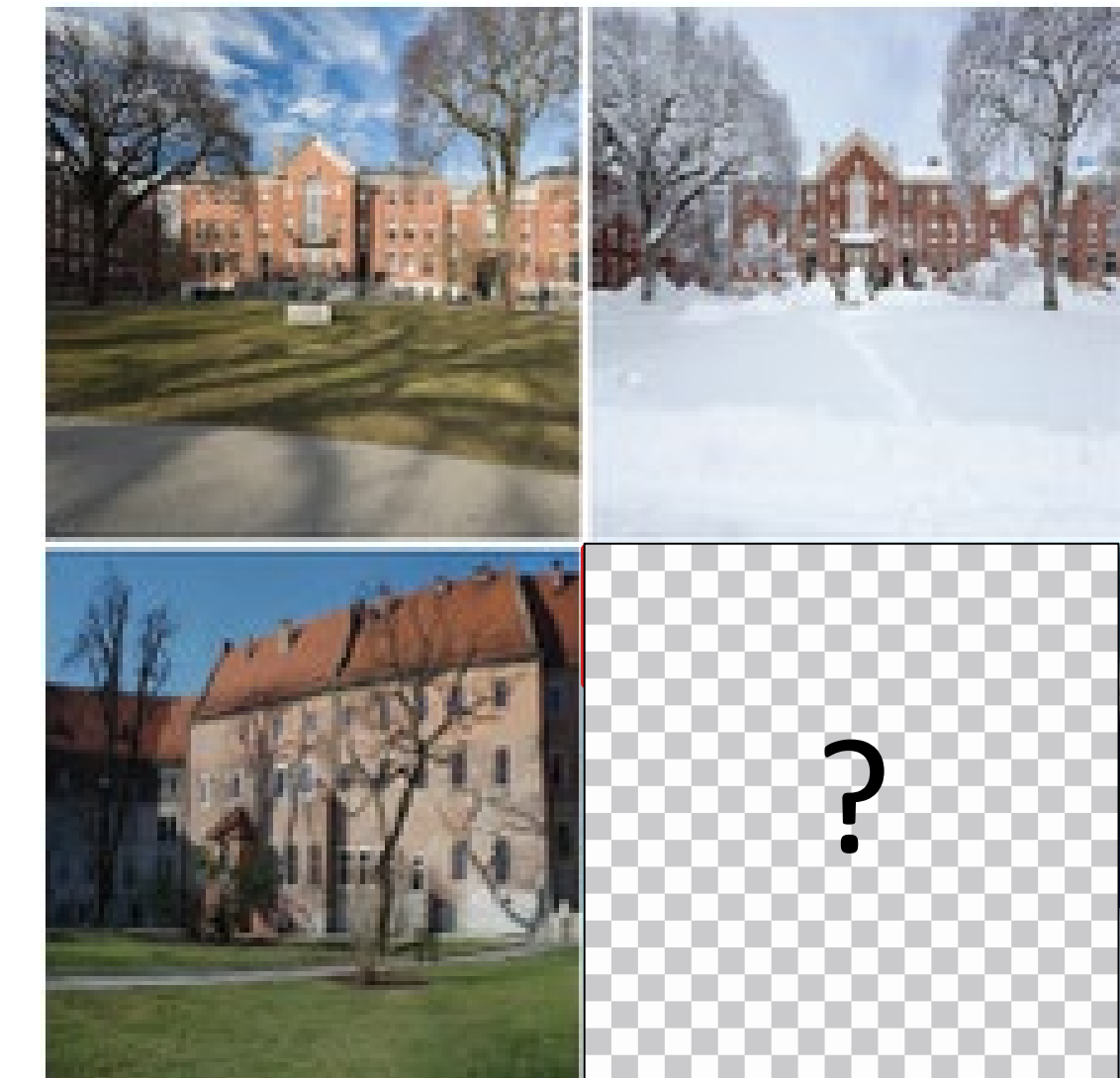
Font Style Transfer



Inpainting



Edge Detection



Style Transfer

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

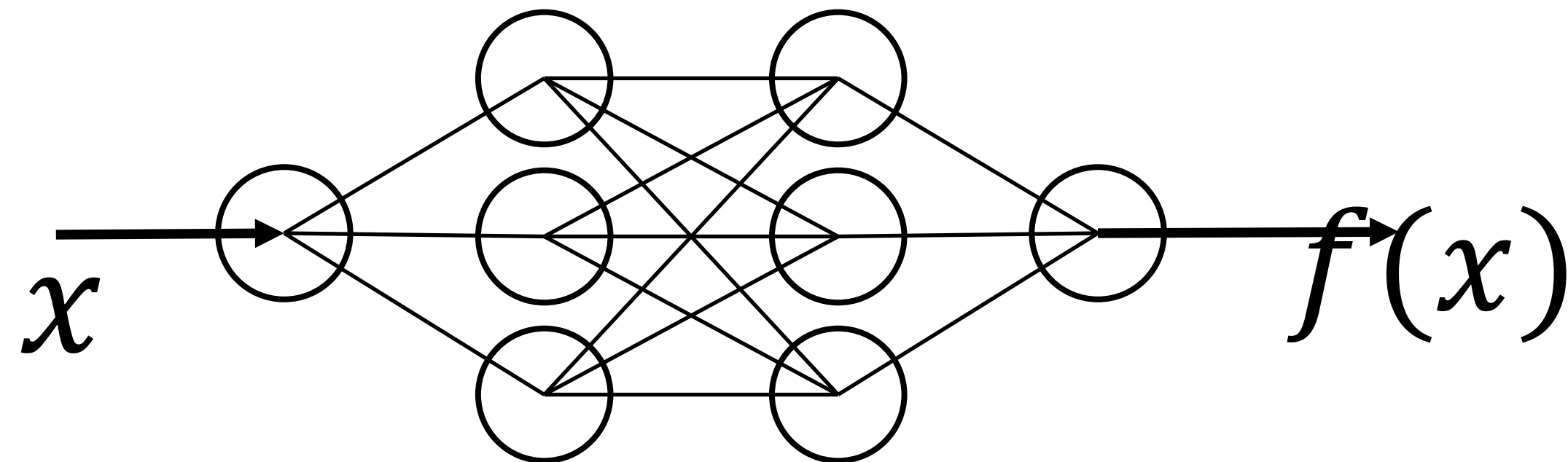
Fourier Series:

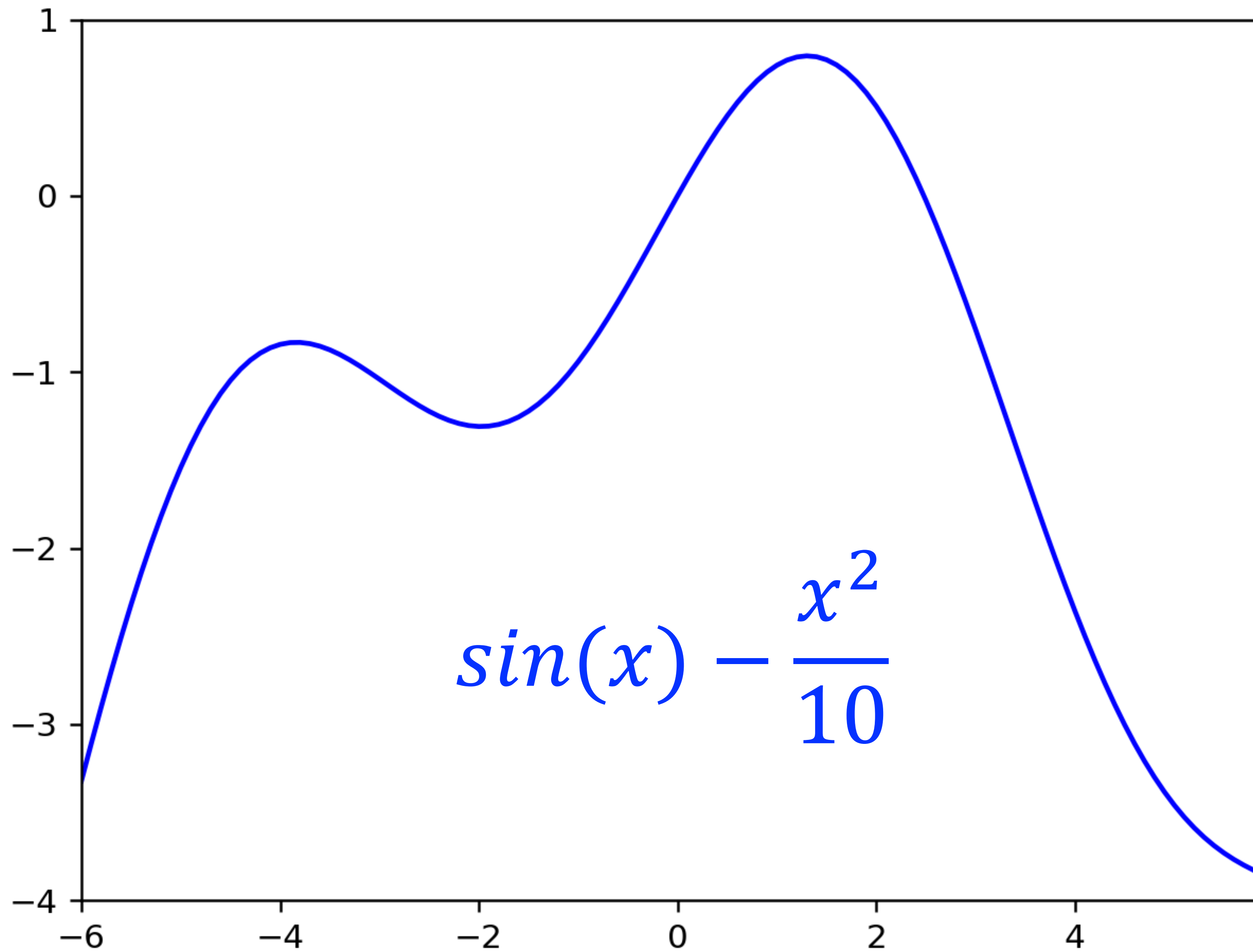
$$f(x) = \text{~} + \text{~} + \text{~} + \text{~} + \dots$$

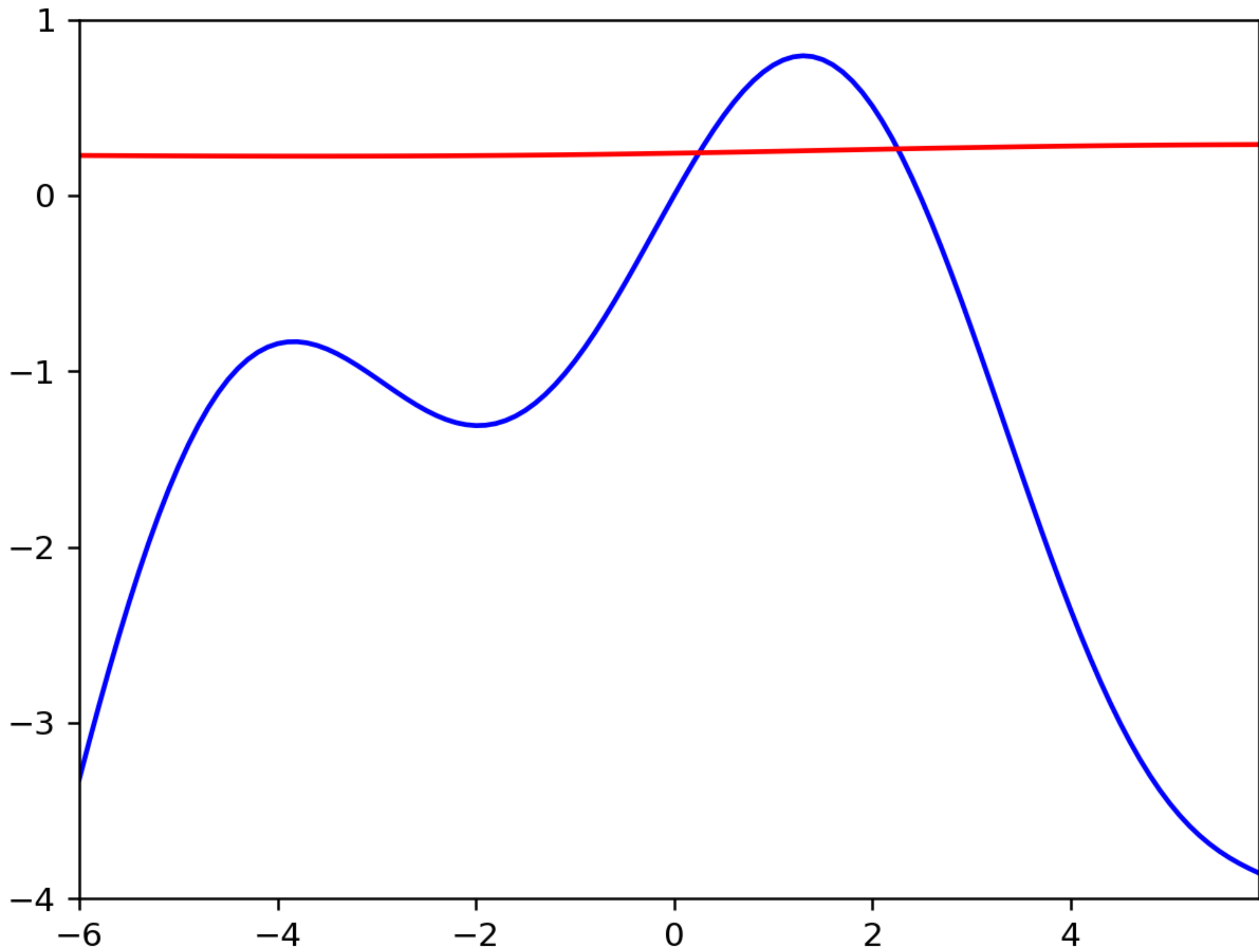
Taylor Series:

$$f(x) = \text{—} + \text{ / } + \text{ U } + \text{ ~ } + \dots$$

Neural Network:







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



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

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



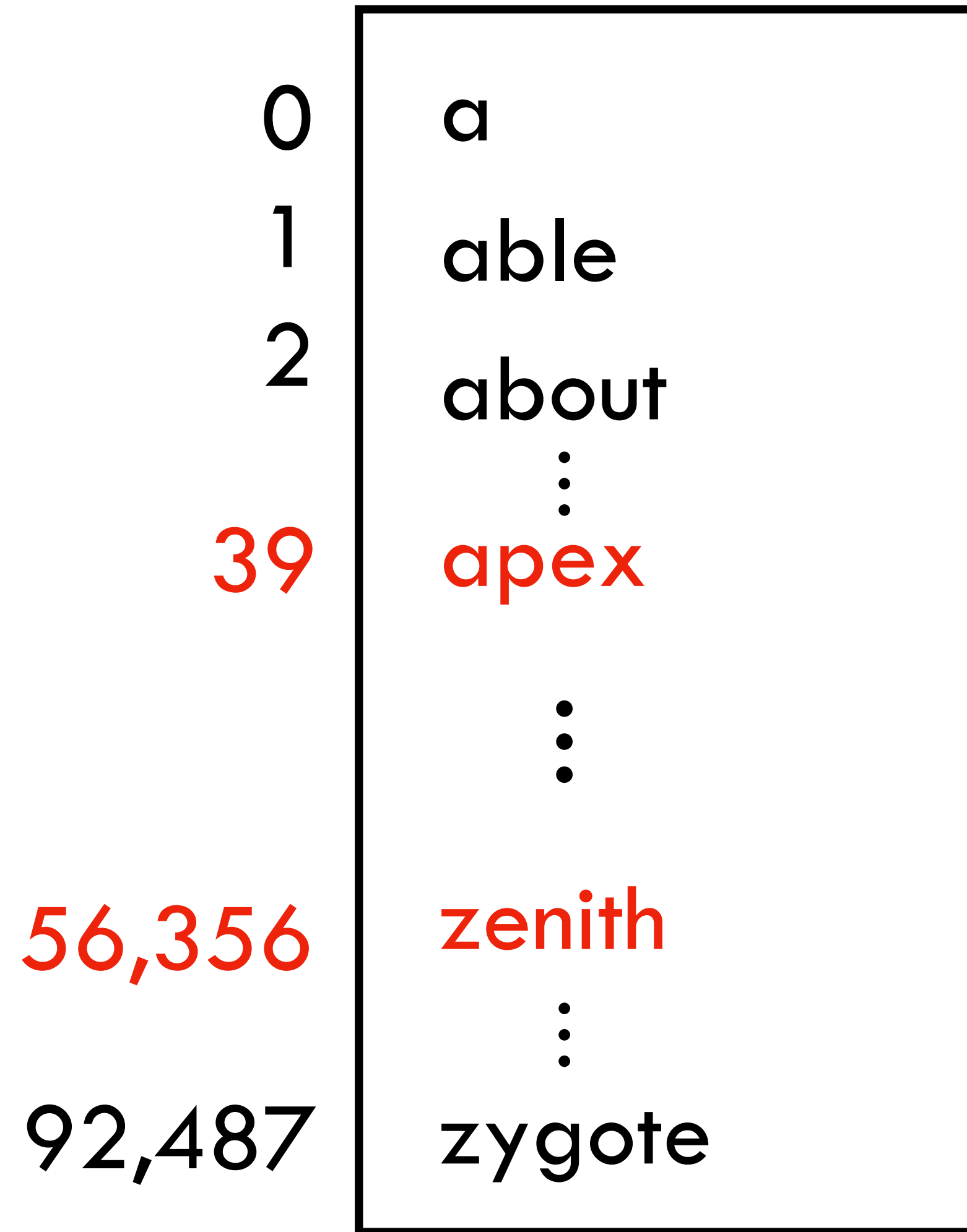
Early one morning the sun was shining I was laying in bed

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

red

neural network

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



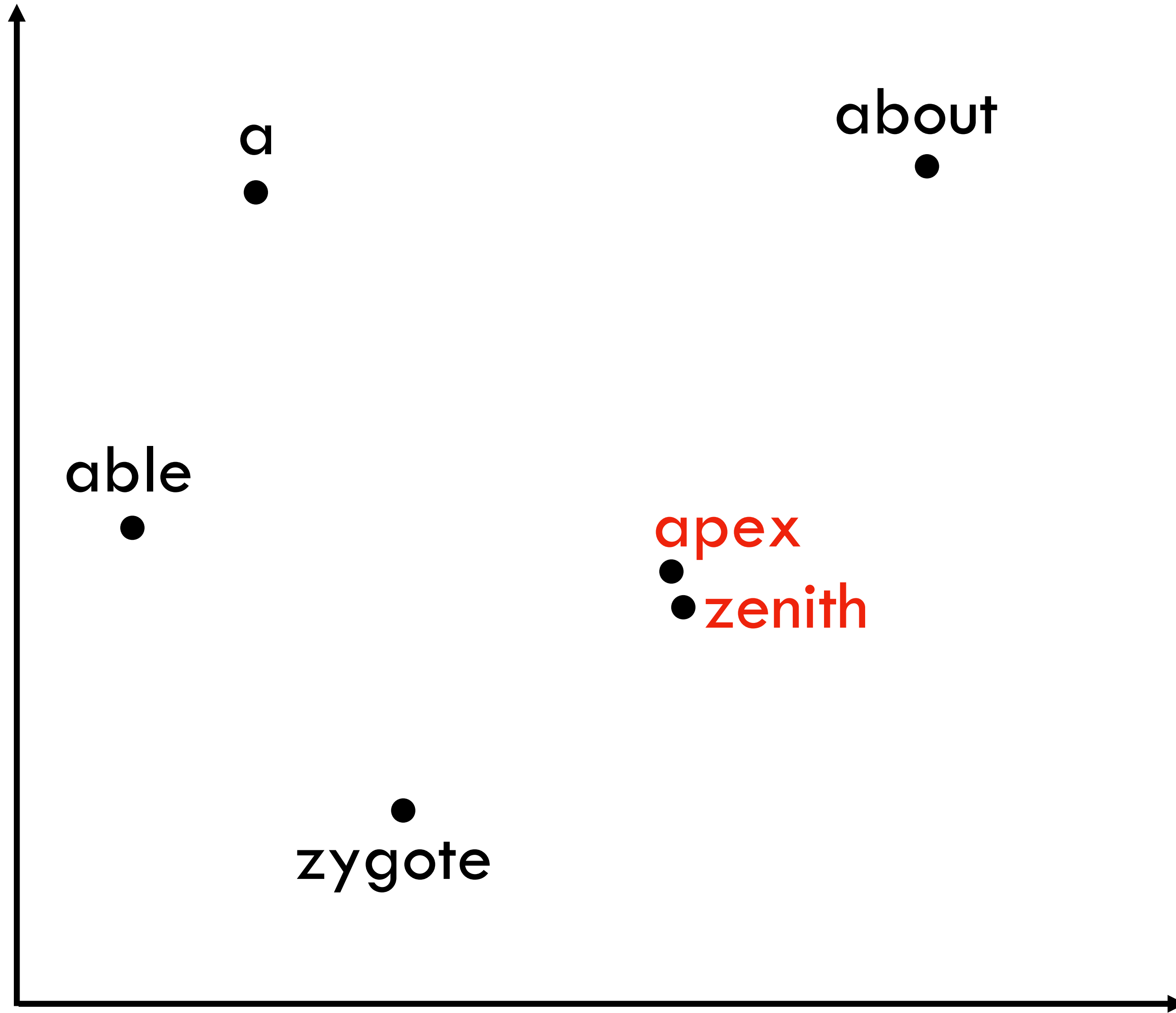
word2vec

[Collobert & Weston 2008; Mikolov et al. 2013]

house, where the professor lived without his wife and child; or so he said jokingly sometimes: "Here's where I live. My house." His daughter often added, without resentment, for the visitor's information, "It started out to be for me, but it's really his." And she might reach in to bring forth an inch-high table lamp with fluted shade, or a blue dish the size of her little fingernail, marked "Kitty" and half full of eternal milk; but she was sure to replace these, after they had been admired, pretty near exactly where they had been. The little house was very orderly, and just big enough for all it contained, though to some tastes the bric-à-brac in the parlor might seem excessive. The daughter's preference was for the store-bought gimmicks and appliances, the toasters and carpet sweepers of Lilliput, but she knew that most adult visitors would



Deep
Net

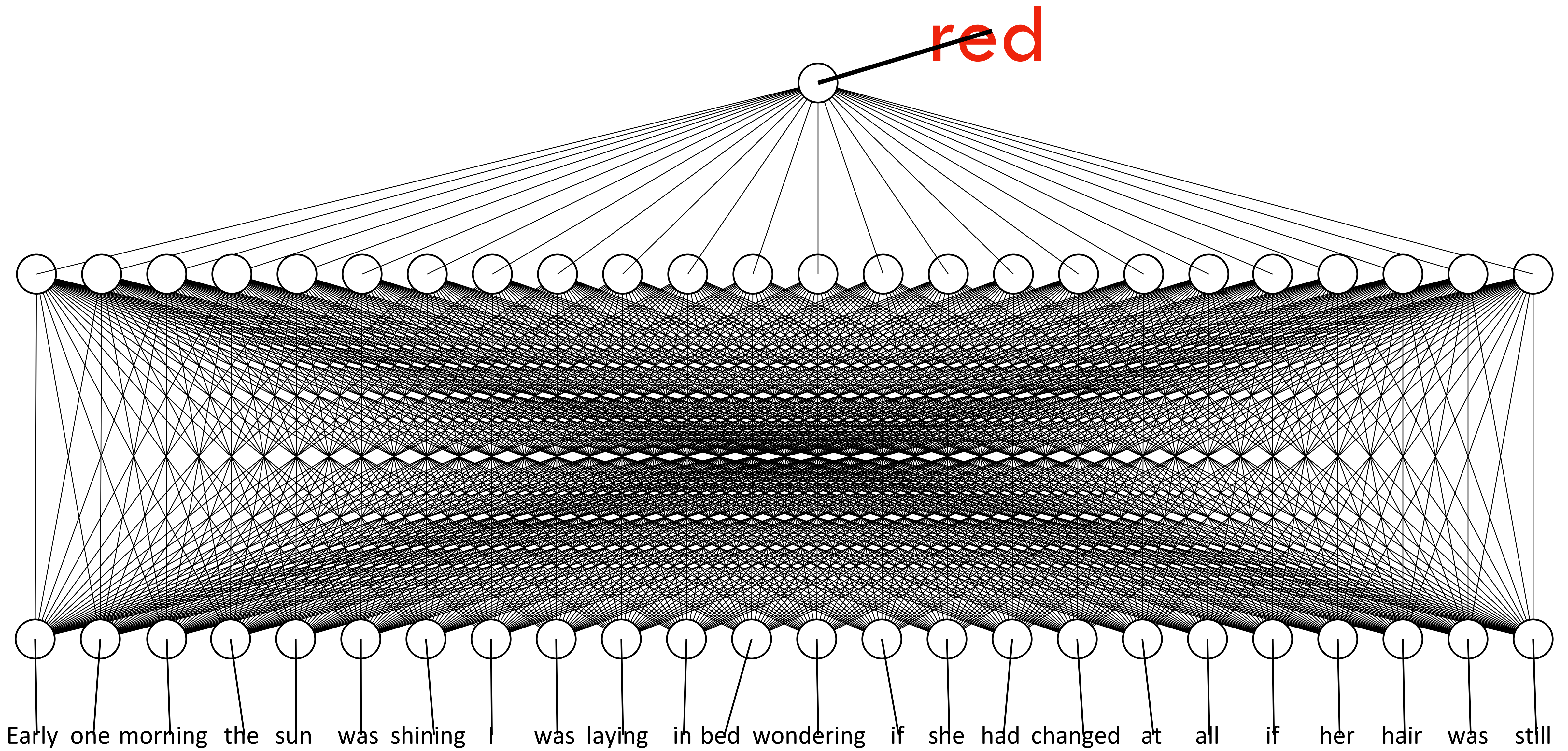


Word Embedding (e.g., word2Vec, GloVe)

red

neural network

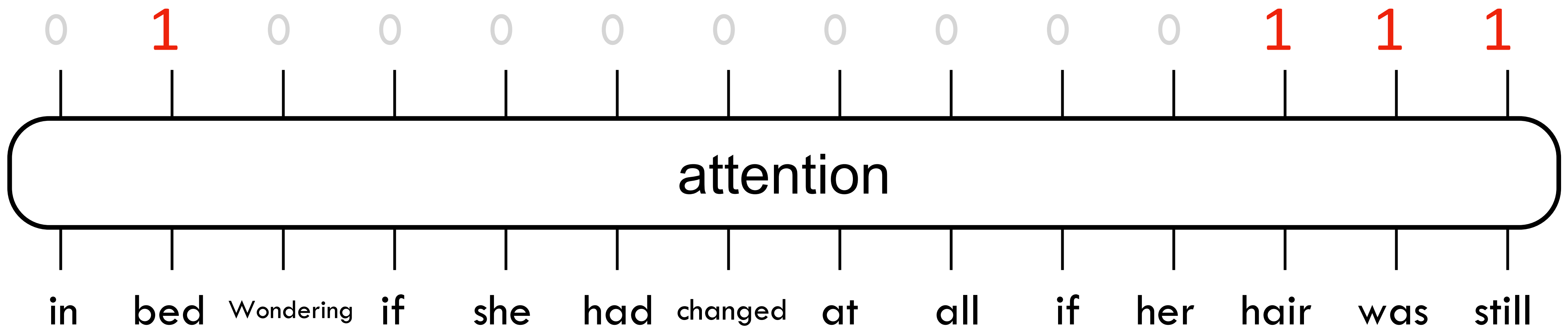
Early one morning the sun was shining I was laying in bed wondering if she had changed at all if her hair was still



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

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

_____ bed
_____ hair was still red



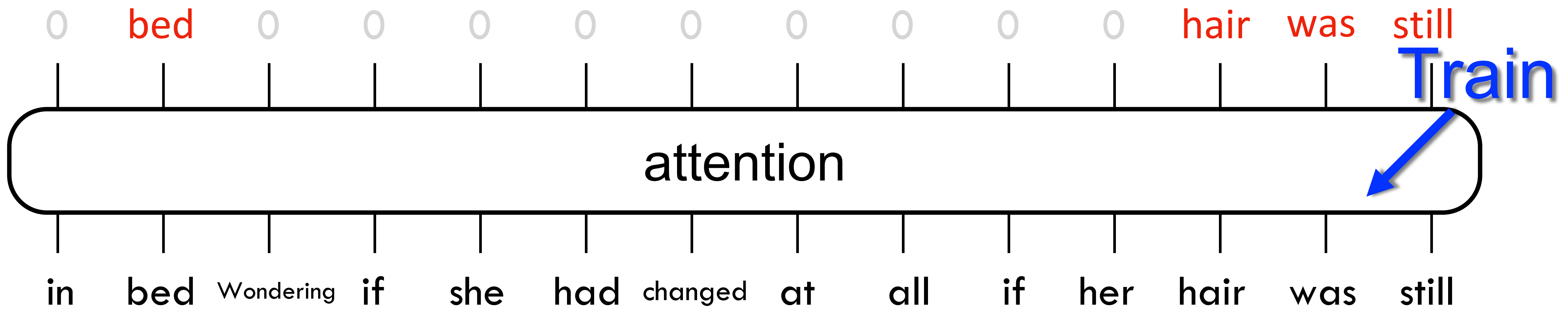
red

next word prediction

0 bed 0 0 0 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



Two roads diverted in a yellow **wood**
And sorry I could not travel **both**
And be one traveler, long I **stood**
And looked down as far as I **could**
To where it bent in the **undergrowth;**

Robert Frost, *Road Not Taken*

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

Train



red

next word prediction

0 bed 0 0 0 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

brown

Train



next word prediction

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

attention

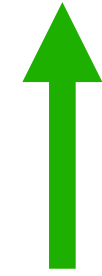
in bed Wondering if she had changed at all if her hair was still

Train

brown

next word prediction

0 bed 0 0 0 0 0 0 0 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

prediction

attention

⋮

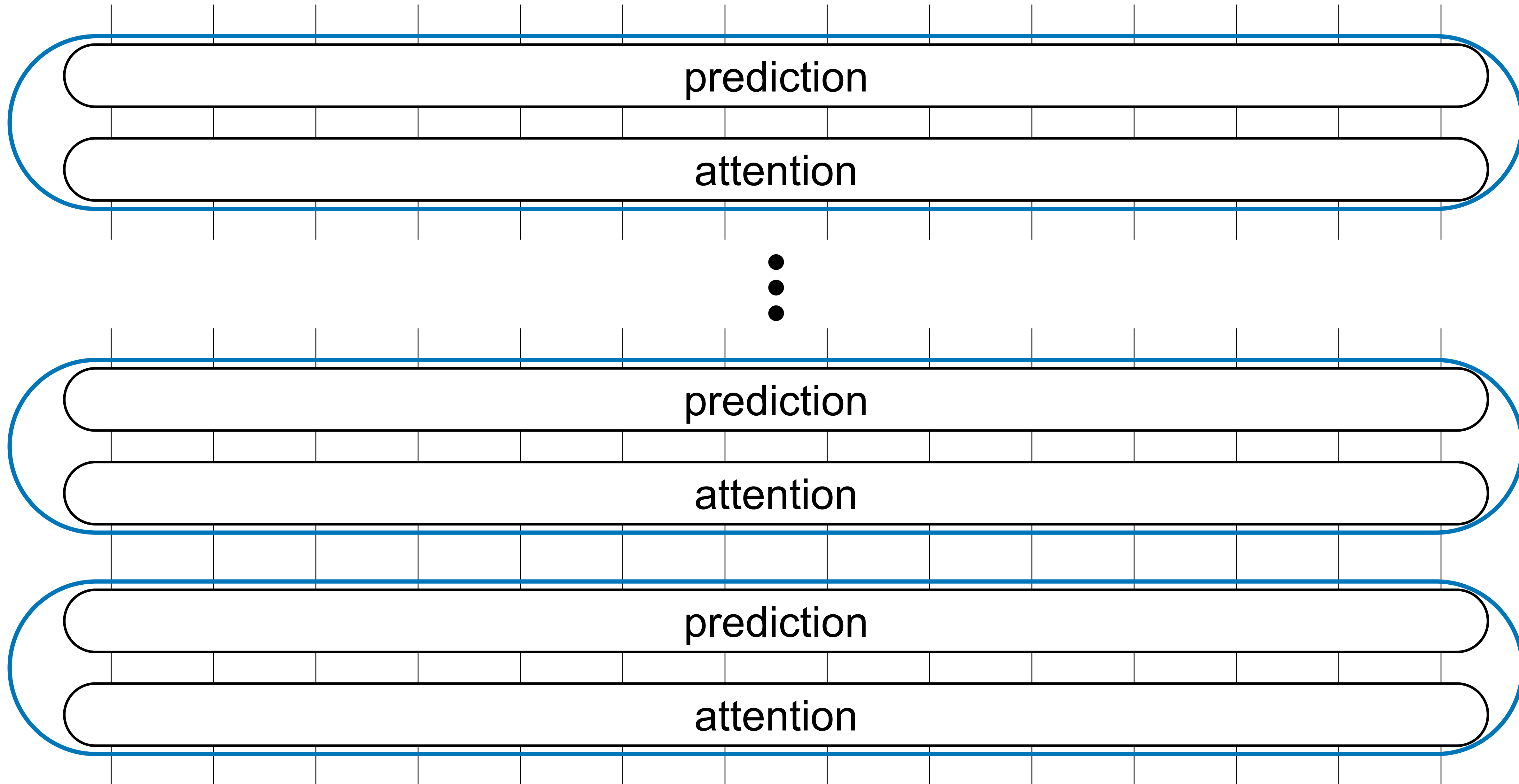
96 (GPT-3) 118 (Palm)

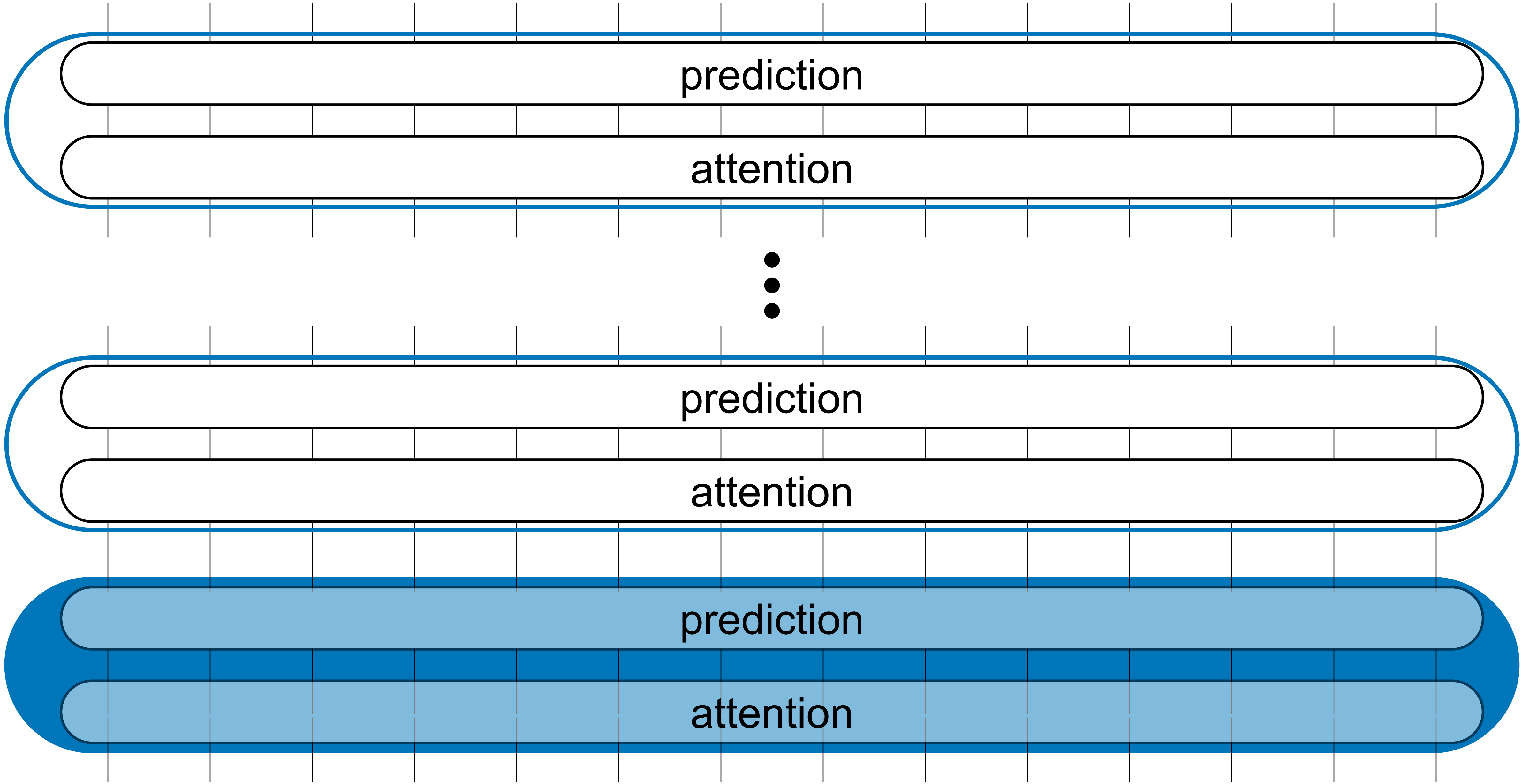
prediction

attention

prediction

attention

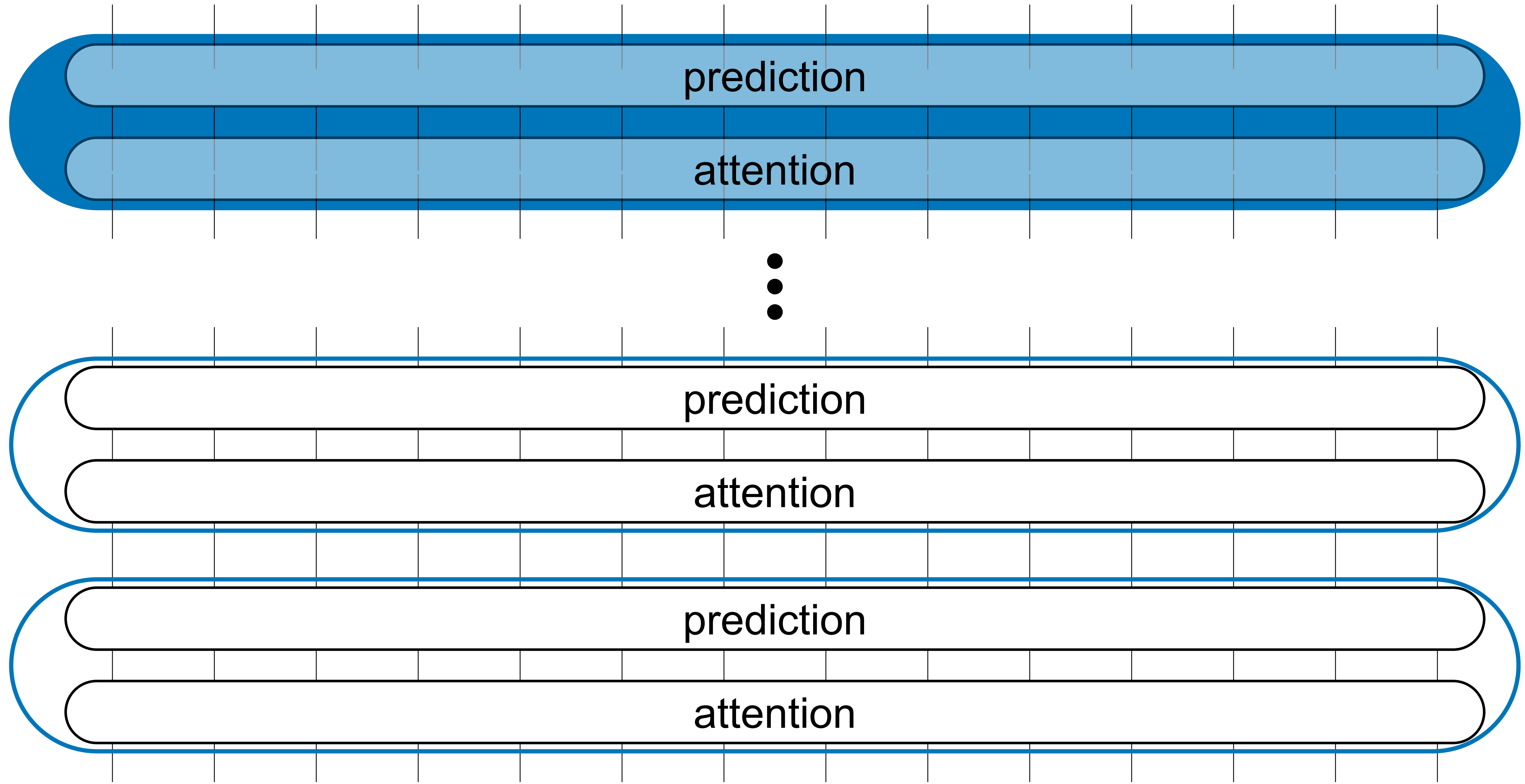


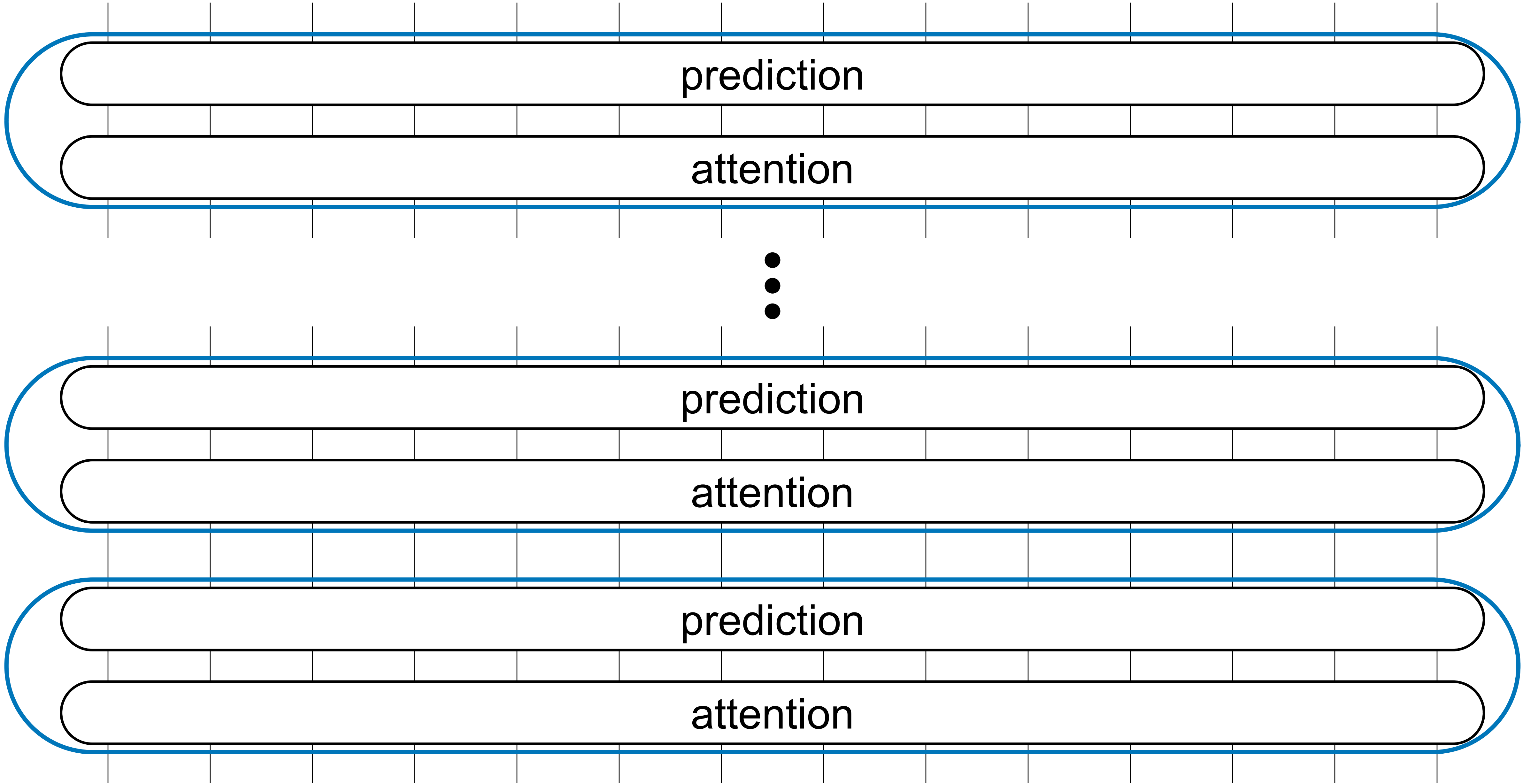


Syntax

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

Semantics





**How much data
to train?**

All of it...

355 years

a month

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

The 16th President was Abraham Lincoln

The capital of Zimbabwe is Harare

Frank Zappa's middle name is Vincent

Napoleon was born on this date 1769

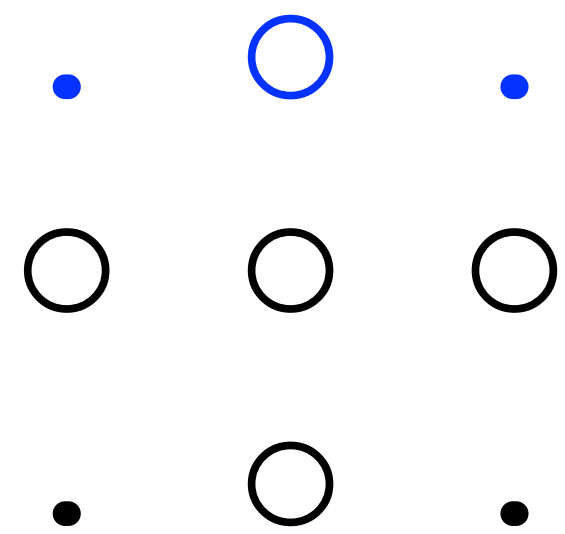
The prime factorization of 19456721434 is $2 \times 3 \times 3 \times 17$

Queen Victoria's maiden name was Alexandrina Victoria

US per-capita income in 1957 was \$2,974

The lat long coordinates of Rome are 41.894722, 12.48

a pattern of characters that looks like a star



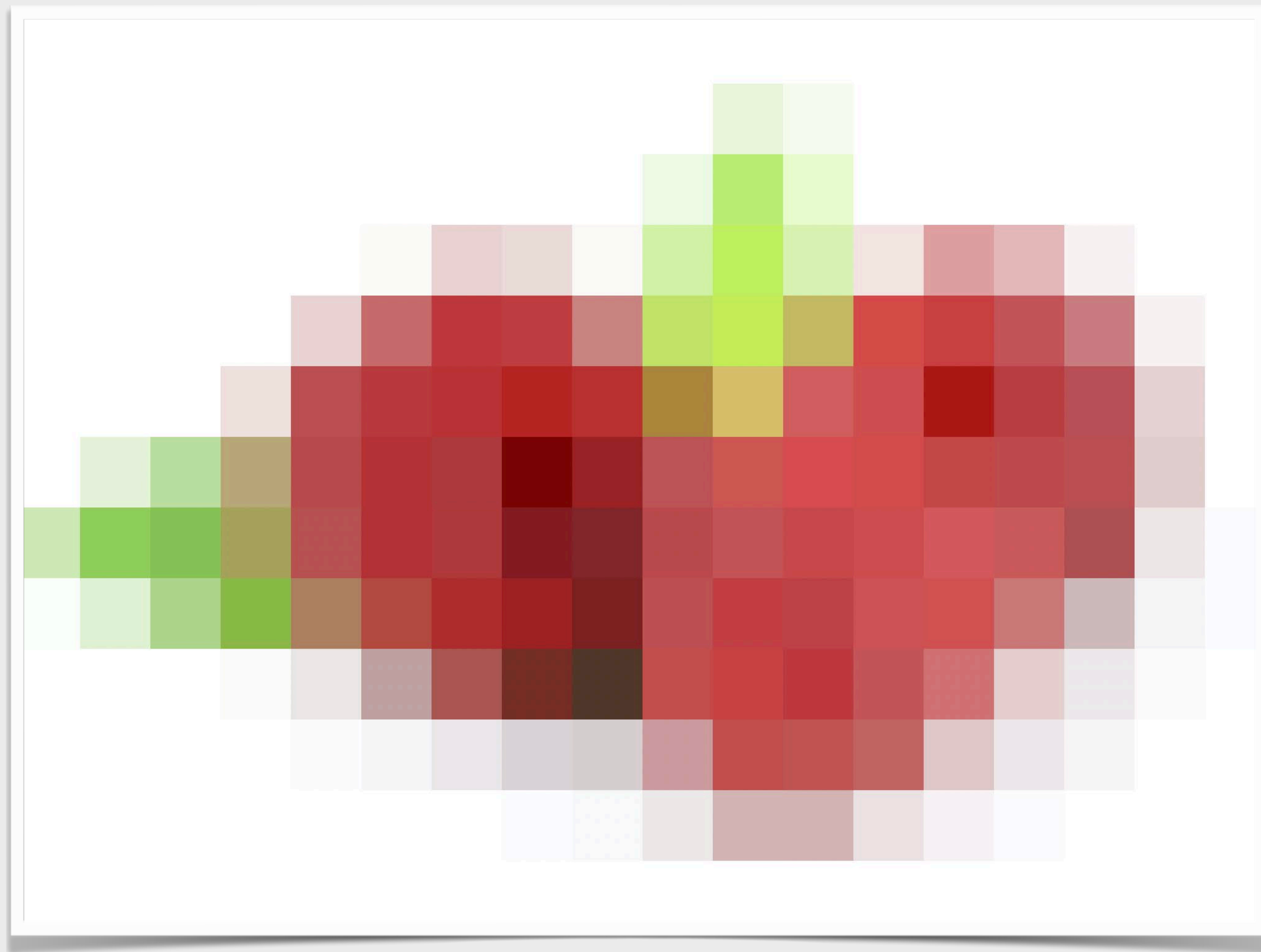
a pattern of characters that looks like a vertical line

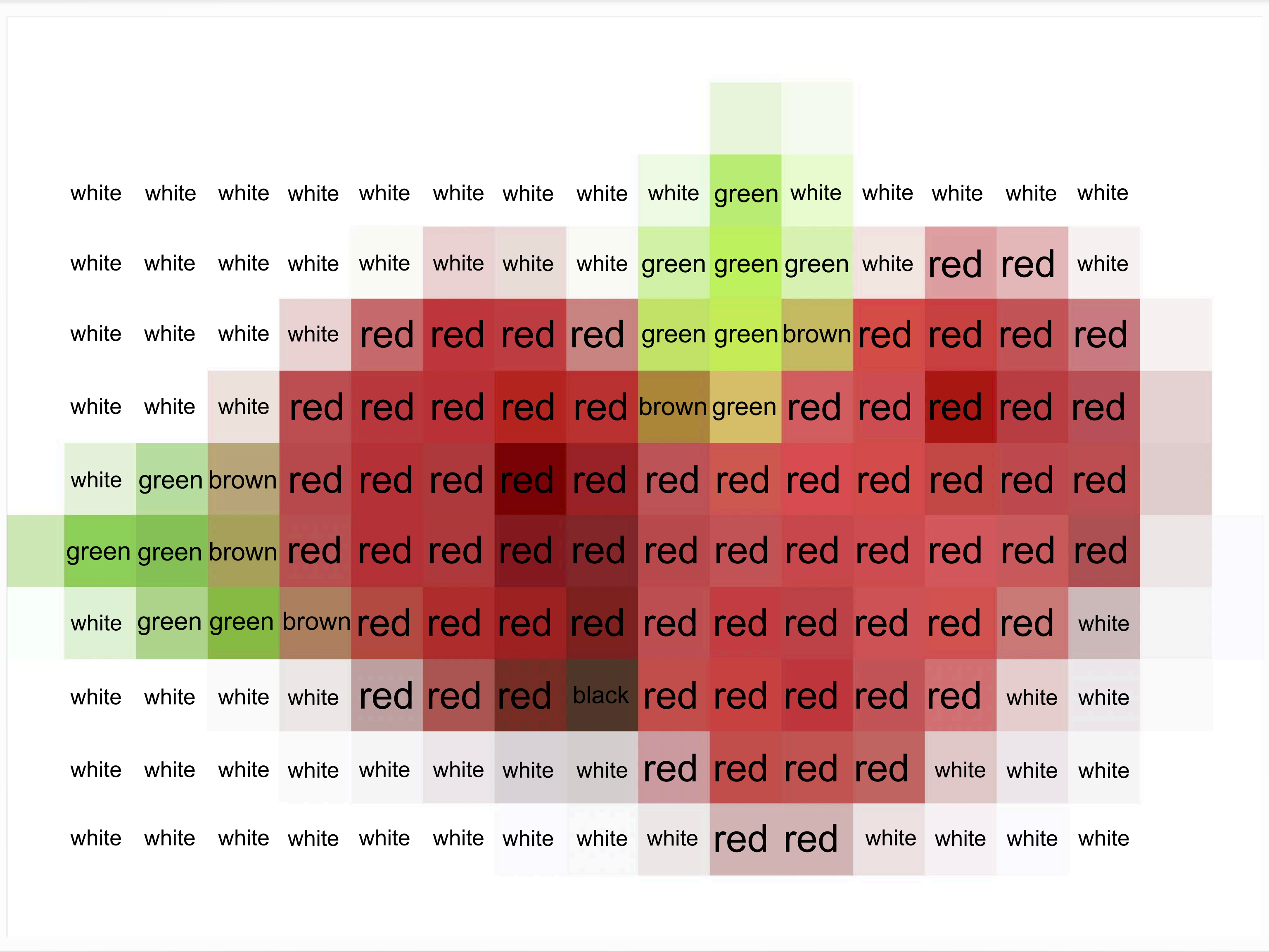
• ○ •
• ○ •
• ○ •
• ○ •
• ○ •

a pattern of characters that looks like a triangle

• ○ •
• ○ ○ •
• ○ ○ ○ •
• ○ ○ ○ ○ •
• ○ ○ ○ ○ ○ •







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

(255,0,0)

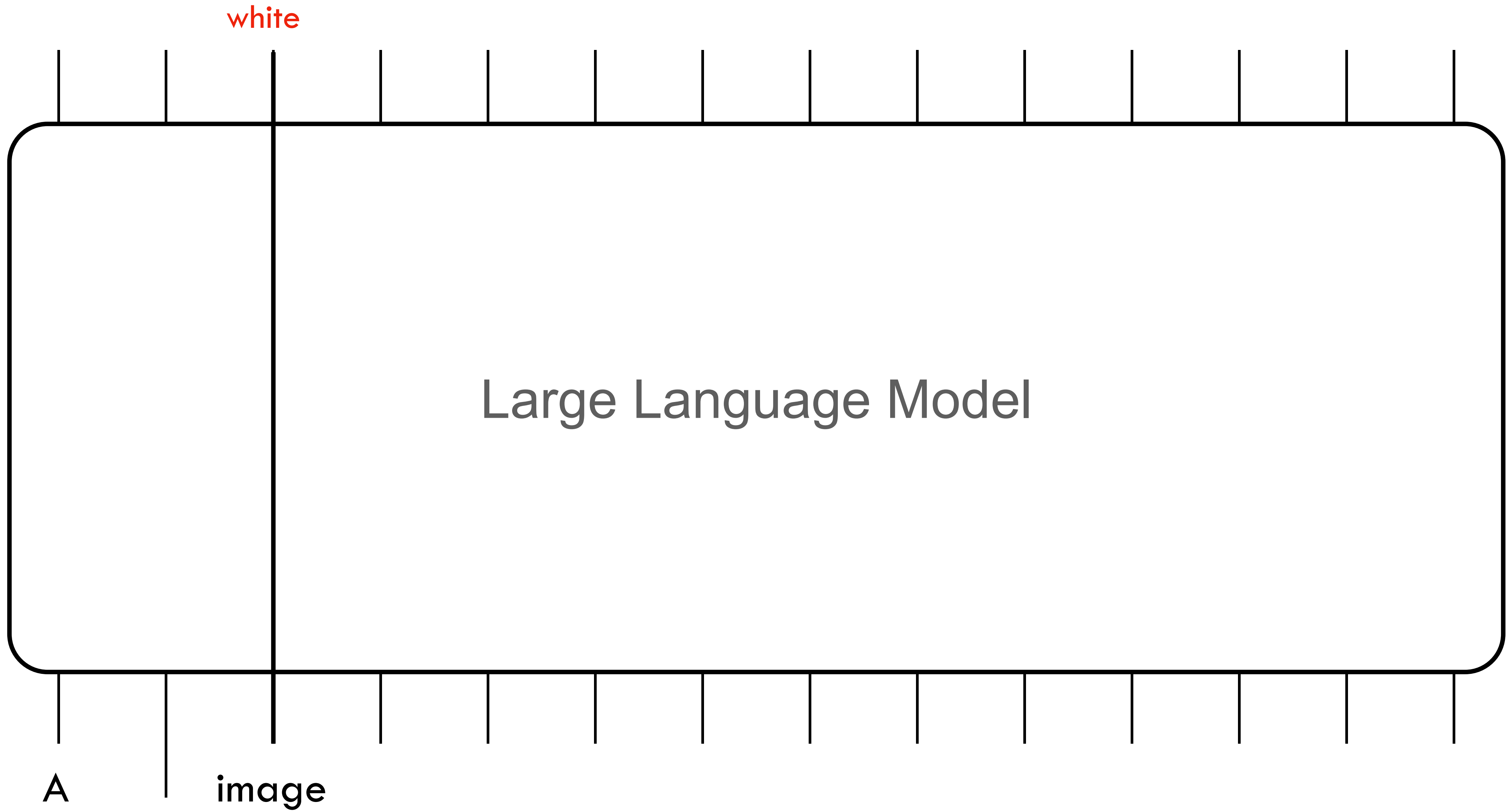
white white white white white white white white white green white white white white white
white white white white white white white white green green green white red red white
white white white white red red red red green green brown red red red red
white white white red red red red red brown green red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



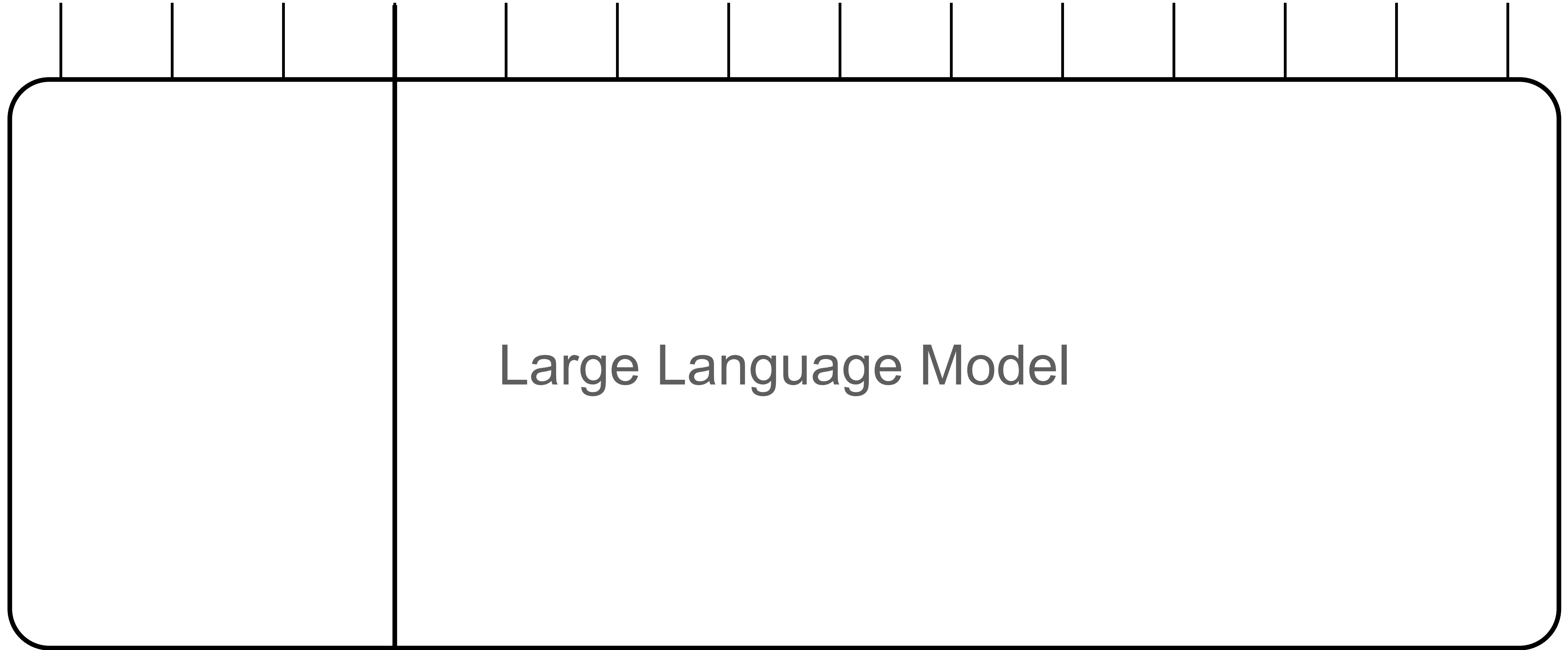
raspberries pancakes sunsets

1 Billion

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



white

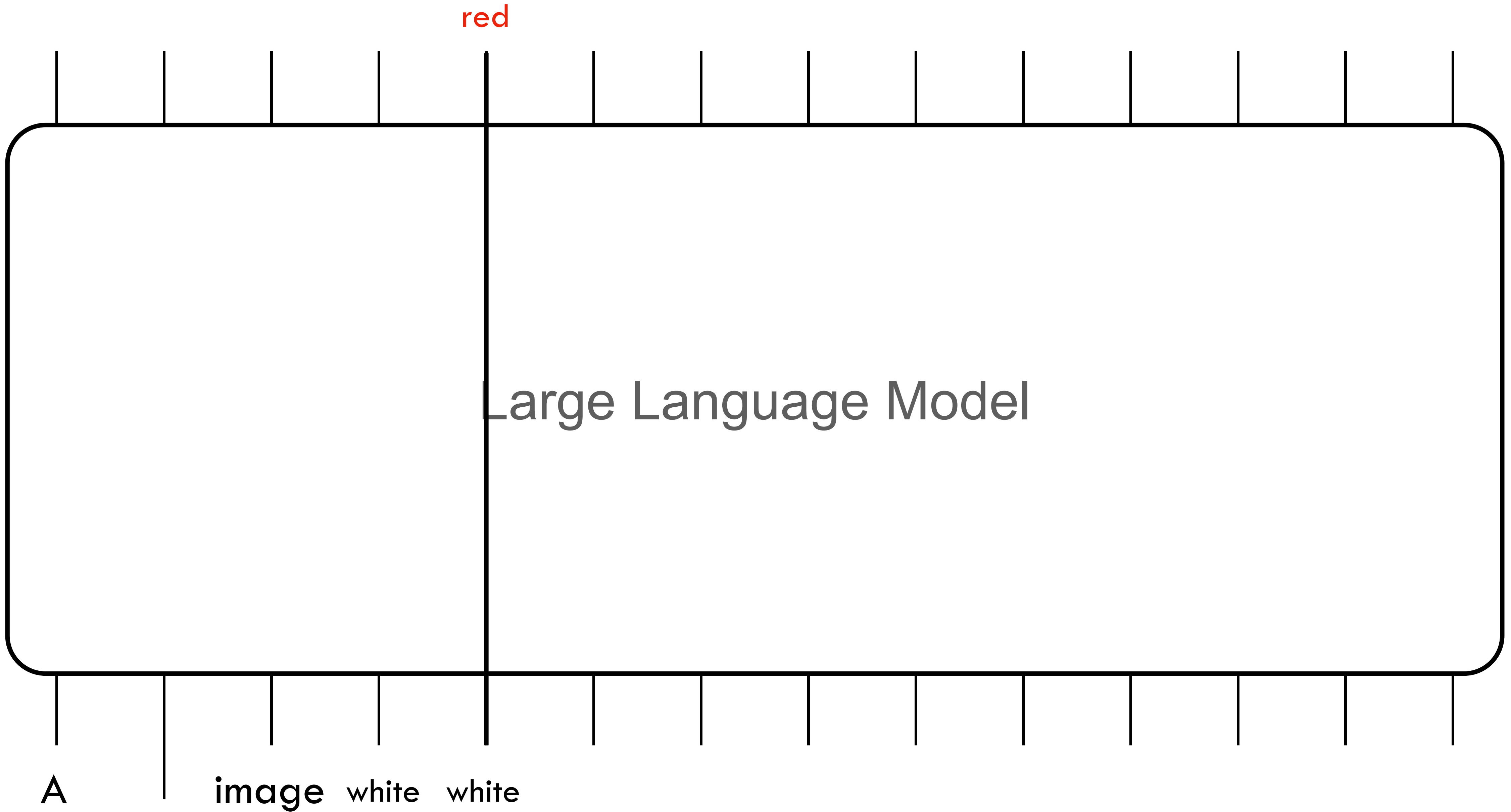


Large Language Model

A

image white

raspberry



Large Language Model

A

image

white

white

red

red

red

white

white

green

green

green

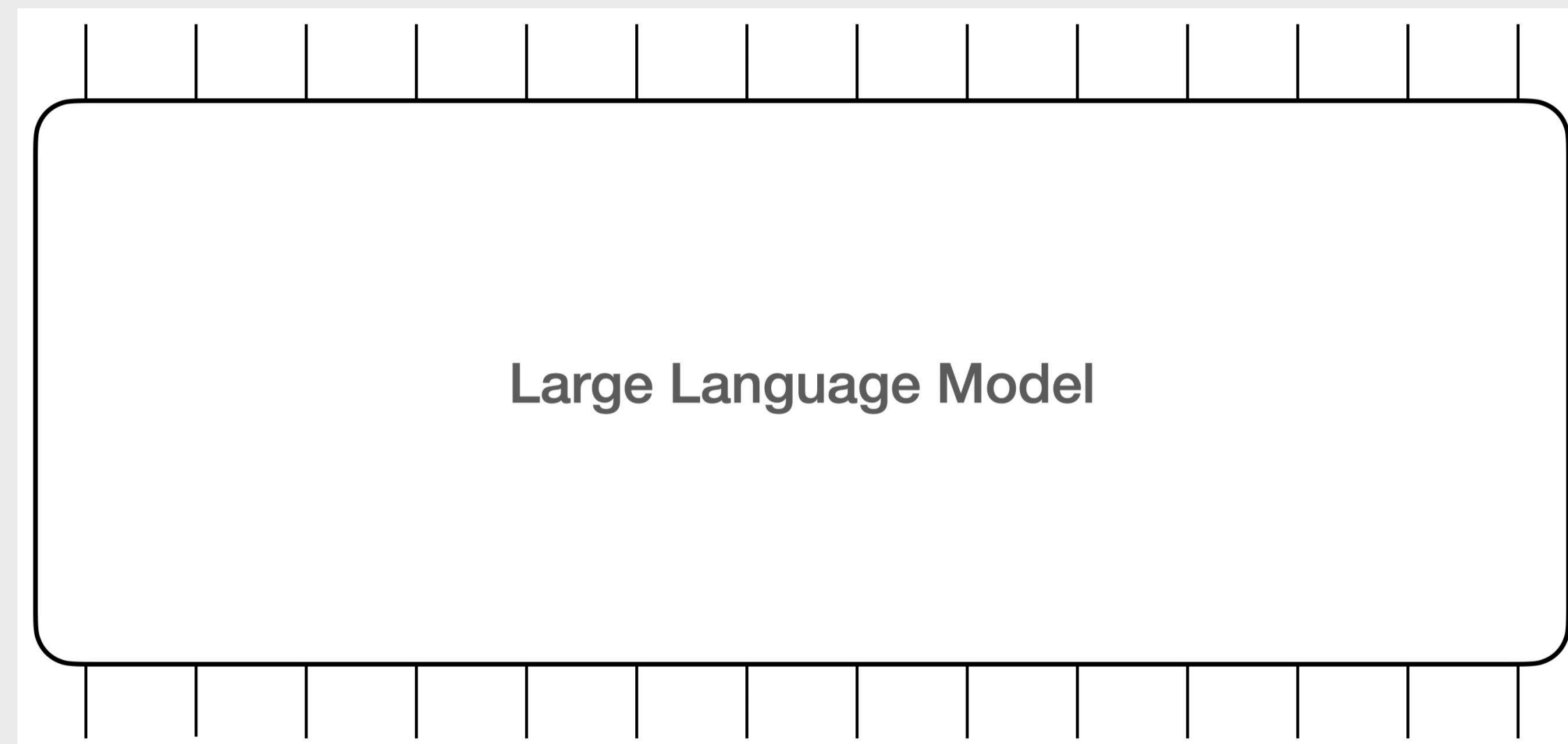
white

raspberry

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

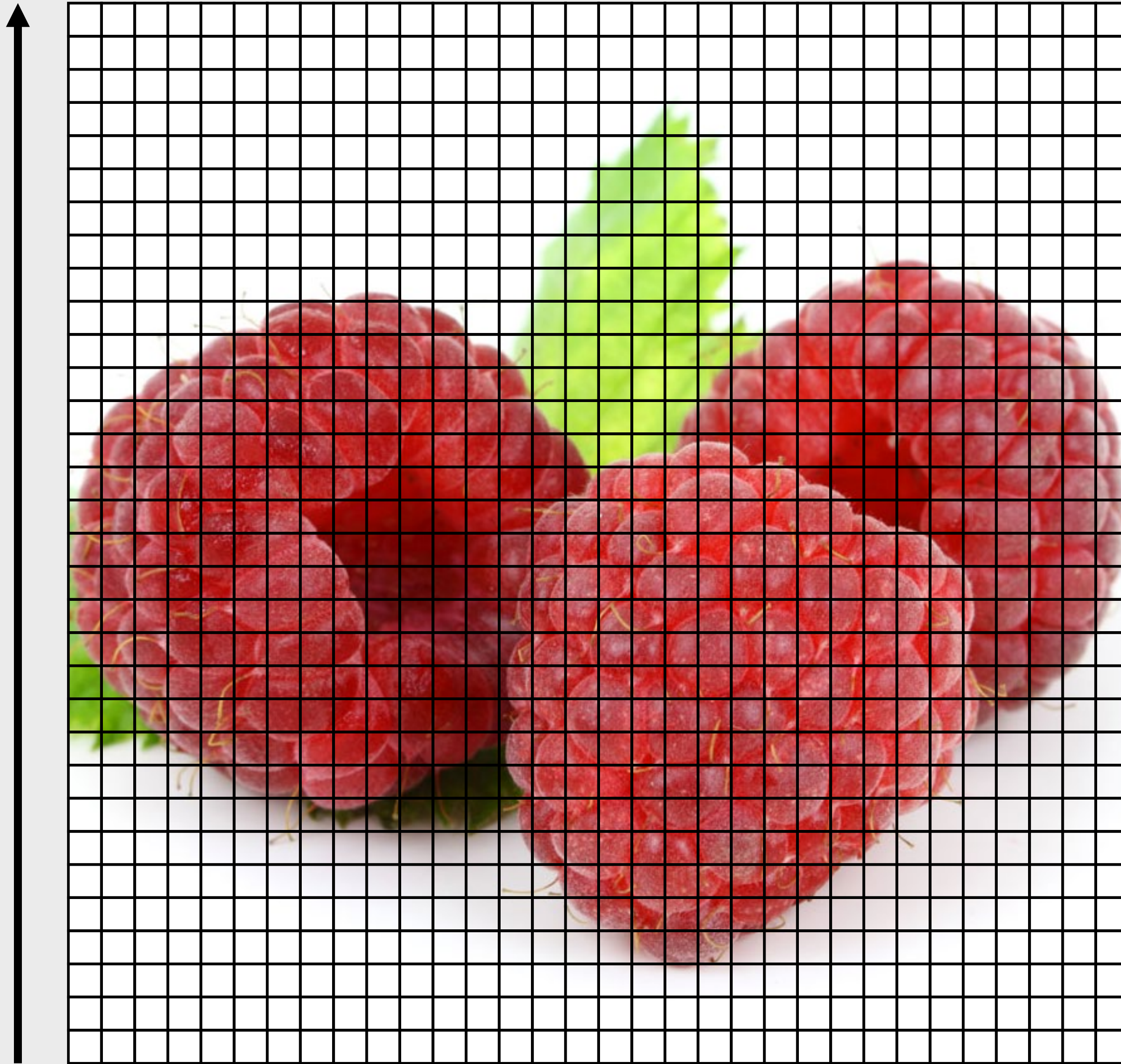


1,000,000s of pixels



1,000s of words

32



32

squirrel reaching for a nut

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1

squirrel reaching for a nut

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1

squirrel reaching for a nut

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6	6	6	1	1	6	6	4	4	4	9	9	9	9	9	9
1	1	1	1	1	7	7	1	1	1	1	1	1	1	1	1	6															

squirrel reaching for a nut

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1

squirrel reaching for a nut

1	1	1	1	1	1	1	1	1	1	1	1	1	1	6	6	6	1	1	6	6	4	4	4	9	9	9	9	9	9
1	1	1	1	1	7	7	1	1	1	1	1	1	1	1	1	6													

squirrel reaching for a nut

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1 6

squirrel reaching for a nut

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1 6

squirrel reaching for a nut

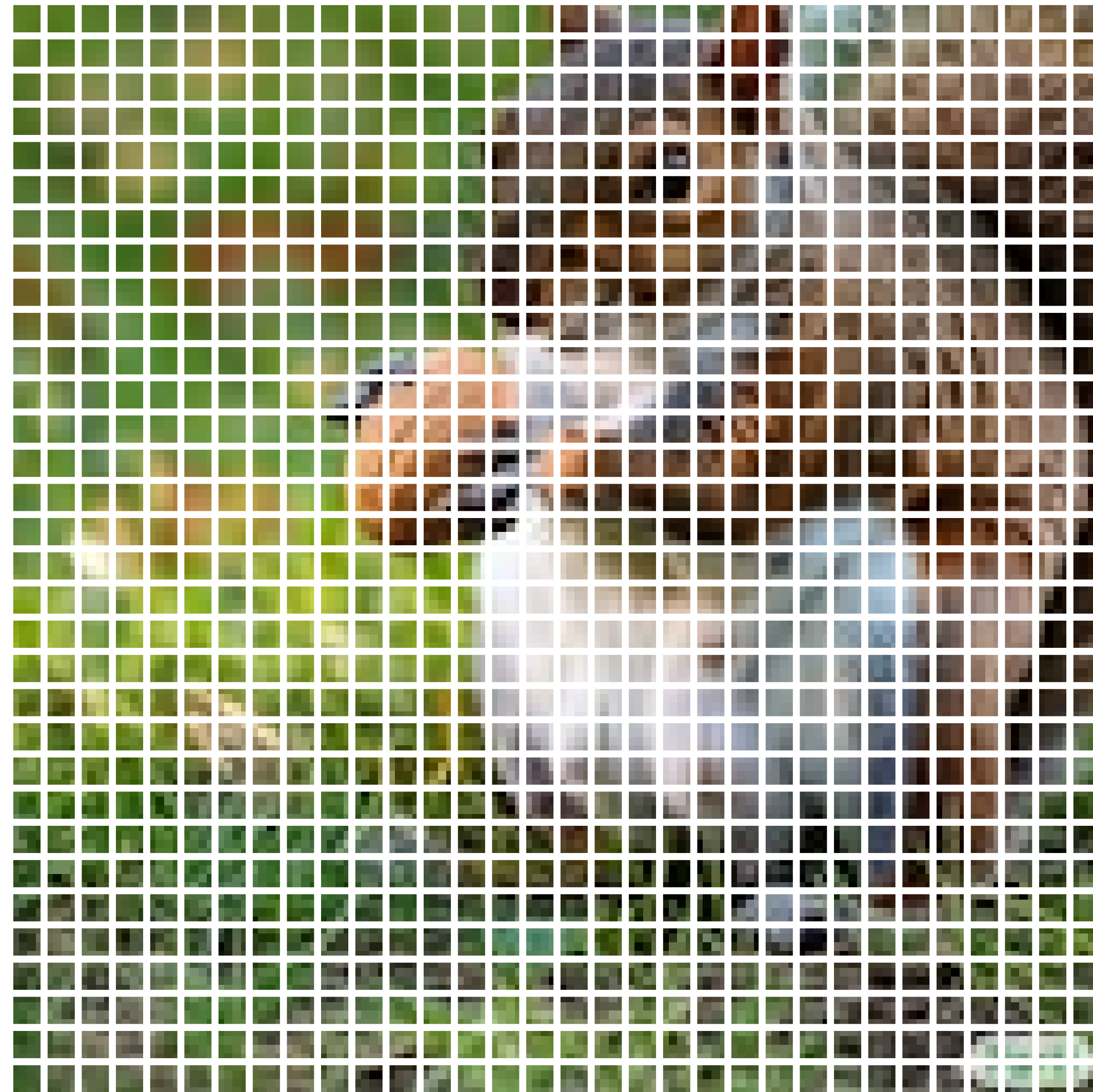
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1 6

squirrel reaching for a nut

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1 6 6 6 6 6 6 6 6 4 1 4 9 9 9 9 9 9
1 1 7 1 1 1 7 1 1 1 1 1 1 1 1 6 6 6 6 6 6 6 6 4 1 4 9 9 9 9 9 9
1 1 7 1 1 1 1 1 1 1 1 1 1 1 2 6 2 6 6 2 2 6 2 9 9 9 9 9 9 9 9
1 1 1 7 7 1 1 1 1 1 1 1 1 1 6 6 6 6 2 5 2 2 4 9 9 9 9 9 9 9 9
1 1 1 7 1 1 1 1 1 1 1 1 1 1 2 6 6 6 2 5 2 2 0 9 9 9 9 9 9 9 9
1 1 1 1 1 1 2 2 2 2 2 1 2 6 6 6 2 6 2 6 2 0 9 9 9 9 9 9 9 9
2 1 1 1 1 2 2 1 2 2 2 1 2 6 6 6 6 6 2 6 4 9 9 9 9 9 9 9 8 8 8
2 2 1 1 1 1 1 1 1 1 1 1 1 6 6 2 6 6 6 4 4 9 9 9 9 9 9 9 8 8
1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 0 0 0 0 0 4 9 9 9 9 9 9 9 8
1 1 1 1 1 1 1 1 1 7 1 5 2 2 2 0 0 0 0 0 0 4 9 9 9 9 9 9 9 8
1 1 1 1 1 1 1 1 1 1 5 2 2 2 2 0 0 0 0 4 4 6 9 9 9 9 9 9 9
1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 5 2 0 5 4 6 9 9 9 9 9 9 9 9
1 1 1 1 7 1 1 1 1 1 2 2 6 6 5 5 2 9 9 9 9 9 9 9 9 9 9 9 9
1 1 1 1 7 2 2 2 1 1 2 2 6 5 5 0 2 9 9 9 9 9 9 9 9 9 9 9 9
1 1 1 7 2 2 2 2 1 1 2 6 6 5 5 0 0 4 4 4 9 9 0 0 0 0 9 9 9 9
1 1 0 1 2 1 2 1 1 1 1 1 1 1 0 0 0 0 4 0 0 0 4 0 0 0 0 9 9 9 8
1 1 1 1 1 1 1 1 1 1 1 7 1 1 0 0 0 0 0 0 0 0 0 0 0 0 9 9 9 8
1 1 1 1 1 1 1 1 1 7 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 9 9 9 8
1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 4 4 9 9 9 8
1 1 1 1 1 7 7 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 4 4 9 9 5 5 9
1 1 1 1 1 7 7 7 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 4 4 9 9 5 4 3
1 1 1 1 1 1 7 7 3 3 3 3 3 3 3 4 0 4 0 0 4 0 0 0 4 4 9 9 5 4 4
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 5 0 0 0 0 0 0 0 4 9 9 9 4 3 3
3 3 3 3 3 3 3 3 3 3 3 4 4 4 3 3 3 5 3 0 4 4 4 4 4 4 9 9 9 4 3 8
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 8 3 4 4 4 4 4 9 3 9 4 8 8
3 3 8 3 3 3 3 3 3 3 3 3 3 3 8 3 8 8 3 3 8 3 8 4 4 8 4 8 3 8 3 3 3
8 8 3 8 8 3 8 3 3 3 8 3 8 3 3 3 3 3 8 3 8 4 8 8 8 3 3 3 3 3 3
3 3 8 3 3 3 3 3 3 3 8 8 8 8 3 3 8 8 3 3 8 8 8 8 8 8 3 8 3 8
3 3 8 8 3 3 3 3 8 3 3 3 3 3 3 3 8 8 3 3 8 3 3 3 8 3 8 8 8 3 8
3 3 8 8 3 3 3 8 8 3 8 3 3 3 3 3 3 3 3 3 3 3 3 3 8 3 8 8 8 8 8
3 3 8 3 3 3 3 8 3 8 8 3 3 3 3 3 3 3 3 3 3 3 3 3 3 8 8 8 8 8 8

squirrel reaching for a nut

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1 6 6 6 6 6 6 6 6 4 1 4 9 9 9 9 9 9
1 1 7 1 1 1 7 1 1 1 1 1 1 1 1 6 6 6 6 6 6 6 6 6 4 1 4 9 9 9 9 9 9
1 1 7 1 1 1 1 1 1 1 1 1 1 1 2 6 2 6 6 2 2 6 2 9 9 9 9 9 9 9 9 9
1 1 1 7 7 1 1 1 1 1 1 1 1 1 6 6 6 6 2 5 2 2 4 9 9 9 9 9 9 9 9 9
1 1 1 7 1 1 1 1 1 1 1 1 1 1 2 6 6 6 2 5 2 2 0 9 9 9 9 9 9 9 9 9
1 1 1 1 1 1 2 2 2 2 2 1 2 6 6 6 2 6 2 6 2 0 9 9 9 9 9 9 9 9 9
2 1 1 1 1 2 2 1 2 2 2 1 2 6 6 6 6 6 2 6 4 9 9 9 9 9 9 9 8 8 8
2 2 1 1 1 1 1 1 1 1 1 1 1 6 6 2 6 6 6 4 4 9 9 9 9 9 9 9 8 8
1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 0 0 0 0 0 4 9 9 9 9 9 9 9 9 8
1 1 1 1 1 1 1 1 1 7 1 5 2 2 2 0 0 0 0 0 0 4 9 9 9 9 9 9 9 9 8
1 1 1 1 1 1 1 1 1 1 5 2 2 2 2 0 0 0 0 4 4 6 9 9 9 9 9 9 9 9 9
1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 5 2 0 5 4 6 9 9 9 9 9 9 9 9 9 9
1 1 1 1 7 1 1 1 1 1 2 2 6 6 5 5 2 9 9 9 9 9 9 9 9 9 9 9 9 9 9
1 1 1 1 7 2 2 2 1 1 2 2 6 5 5 0 2 9 9 9 9 9 9 9 9 9 9 9 9 9 9
1 1 1 7 2 2 2 2 1 1 2 6 6 5 5 0 0 4 4 4 9 9 0 0 0 0 9 9 9 9 9
1 1 0 1 2 1 2 1 1 1 1 1 1 1 0 0 0 0 4 0 0 0 4 0 0 0 0 9 9 9 9 8
1 1 1 1 1 1 1 1 1 1 1 7 1 1 0 0 0 0 0 0 0 0 0 0 0 0 9 9 9 9 8
1 1 1 1 1 1 1 1 1 7 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 9 9 9 8 8
1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 4 4 9 9 9 8 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 4 4 9 9 5 5 9
1 1 1 1 1 7 7 7 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 4 4 9 9 5 4 3
1 1 1 1 1 1 7 7 3 3 3 3 3 3 3 4 0 4 0 0 4 0 0 0 0 4 4 9 9 5 4 4
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 5 0 0 0 0 0 0 0 4 9 9 9 4 3 3
3 3 3 3 3 3 3 3 3 3 3 4 4 4 3 3 3 5 3 0 4 4 4 4 4 4 9 9 9 4 3 8
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 8 3 4 4 4 4 4 9 3 9 4 8 8
3 3 8 3 3 3 3 3 3 3 3 3 3 3 8 3 8 8 3 3 8 3 8 4 4 8 4 8 3 8 3 3 3
8 8 3 8 8 3 8 3 3 3 8 3 8 3 3 3 3 3 8 3 8 4 8 8 8 3 3 3 3 3 3
3 3 8 3 3 3 3 3 3 3 8 8 8 8 3 3 8 8 3 3 8 8 8 8 8 8 3 8 3 8
3 3 8 8 3 3 3 3 8 3 3 3 3 3 3 3 8 8 3 3 8 3 3 3 8 3 8 8 8 3 8
3 3 8 8 3 3 3 8 8 3 8 3 3 3 3 3 3 3 3 3 3 3 3 3 8 3 8 8 8 8 8
3 3 8 3 3 3 3 8 3 8 8 3 3 3 3 3 3 3 3 3 3 3 3 3 3 8 8 8 8 8 8



squirrel reaching for a nut



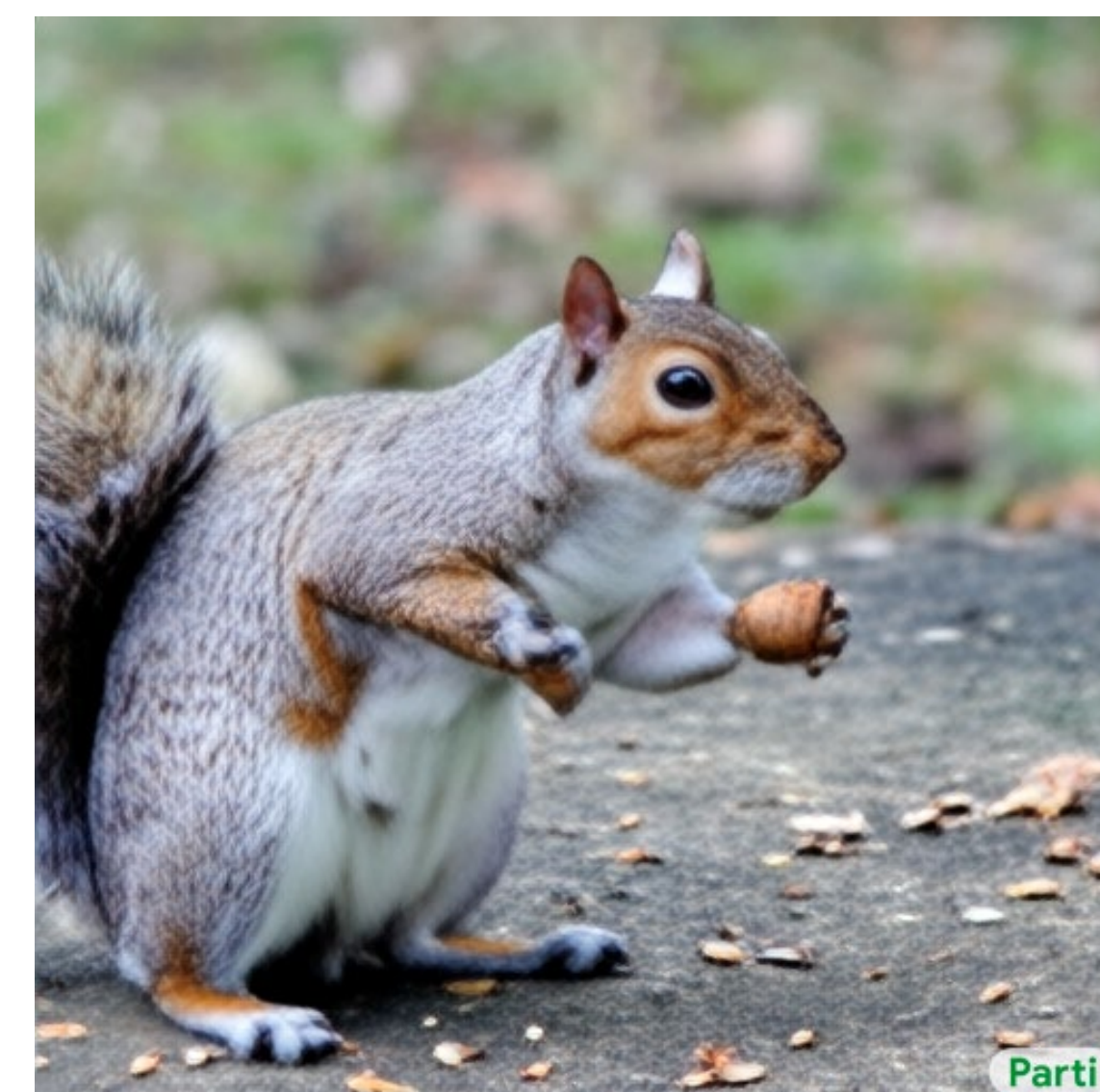
squirrel reaching for a nut



**Up-sampled
4x**

squirrel reaching for a nut

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



squirrel reaching for a nut

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



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](#)



squirrel made of toothpicks wearing sunglasses reaching for a nut

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



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

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



Squirrel reaching for a nut. by Leonardo da Vinci

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



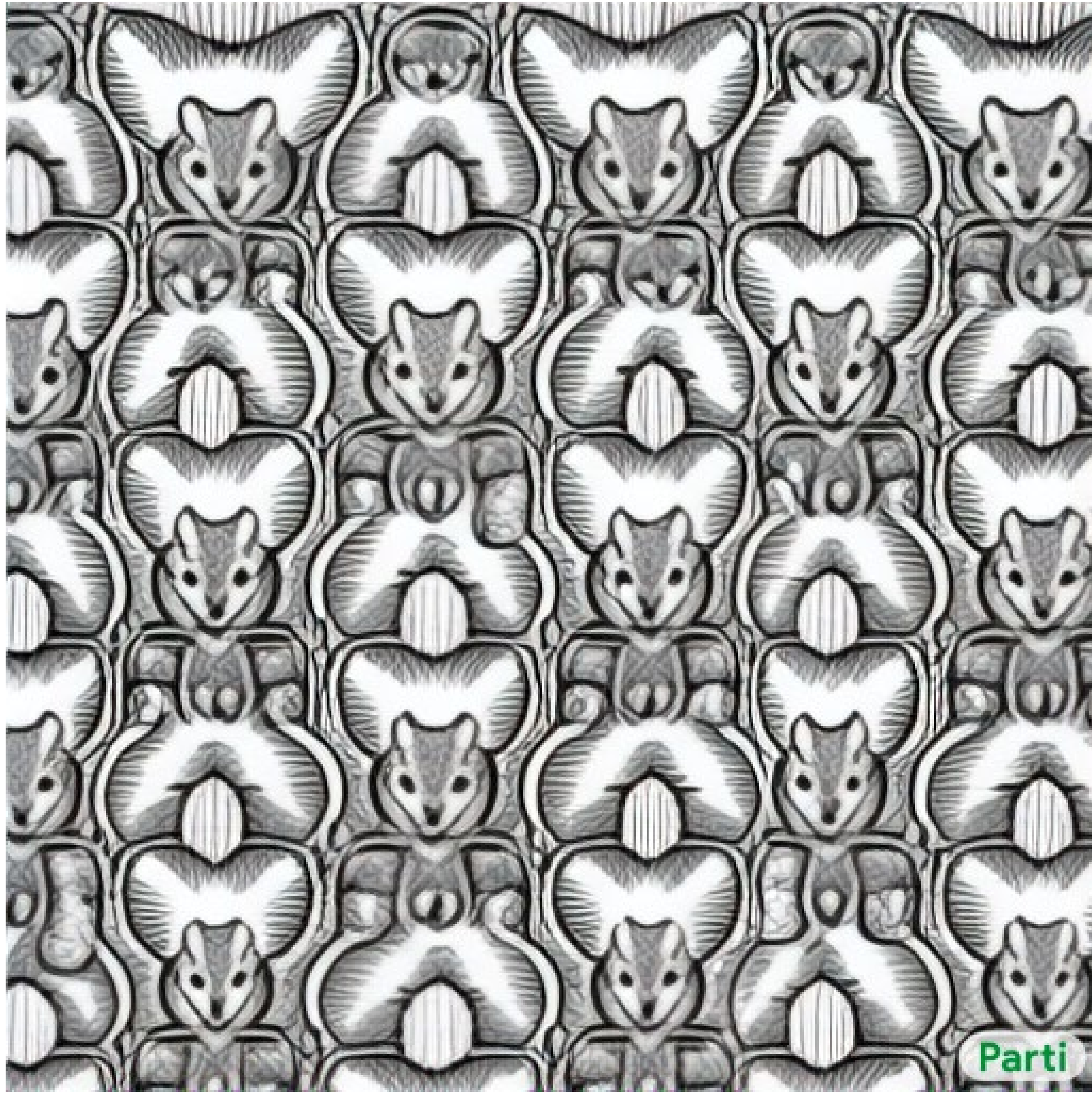
Squirrel reaching for a nut. Van Gogh painting

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



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](#)

Algorithm vs. Data

Diffusion-based

Auto-regressive

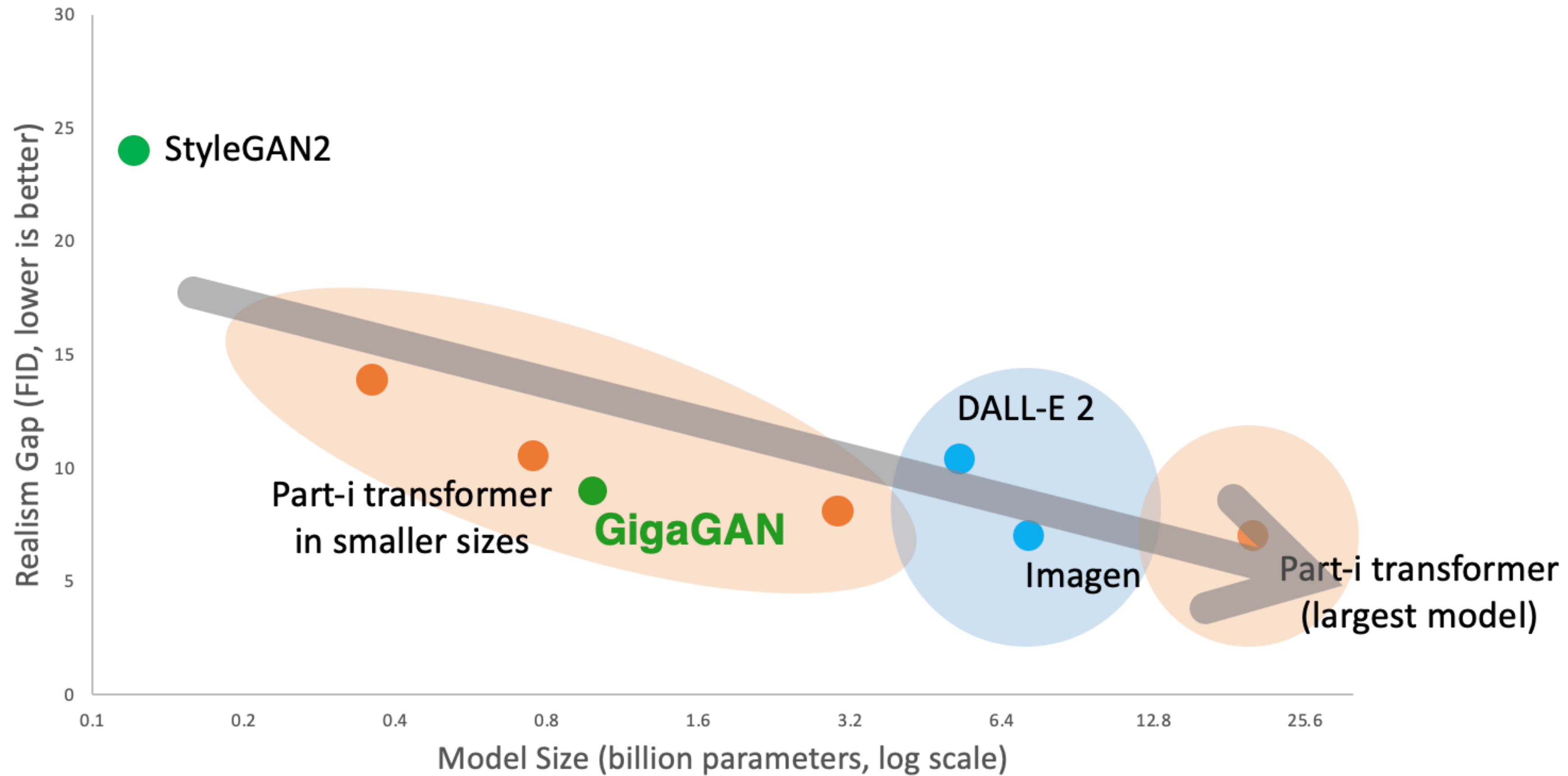
GAN-based



Prompt: *“squirrel reaching for a nut”*

data capacity vs. image quality

Larger Models Attain Better Realism



Graph by Taesung Park

Generative Magic

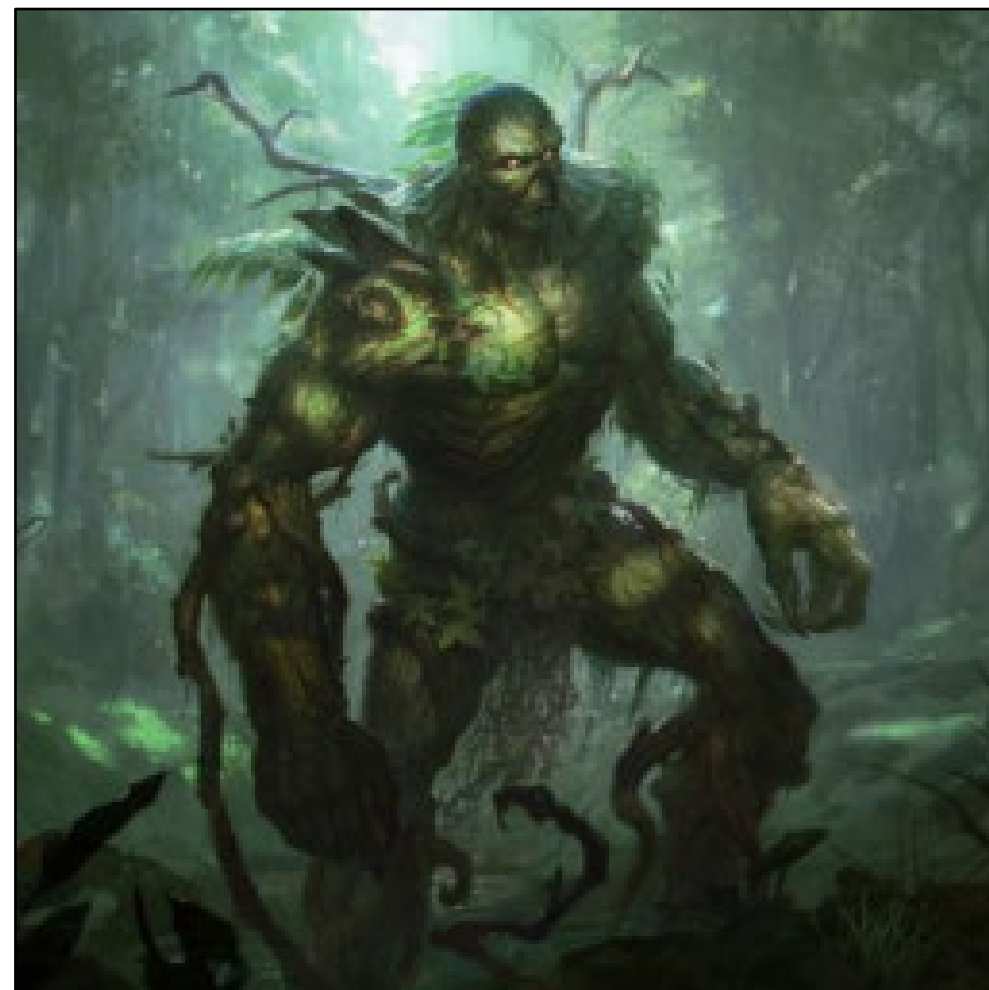


Prompt: “a green creature made of leaves and vines bursting out of the ground ready to attack; detailed, best on artstation, raymond swanland, magic the gathering, epic, stunning, masterpiece”

“influences” from the training data



generated image



to paraphrase Arthur C. Clarke:

*Interpolation in sufficiently
high-dimensional space is
indistinguishable from magic*