



Generative AI with Diffusion Models

Part 5: CLIP

Agenda

- Part 1: From U-Nets to Diffusion

- Part 2: Denoising Diffusion Probabilistic Models

- Part 3: Optimizations

- Part 4: Classifier-Free Diffusion Guidance

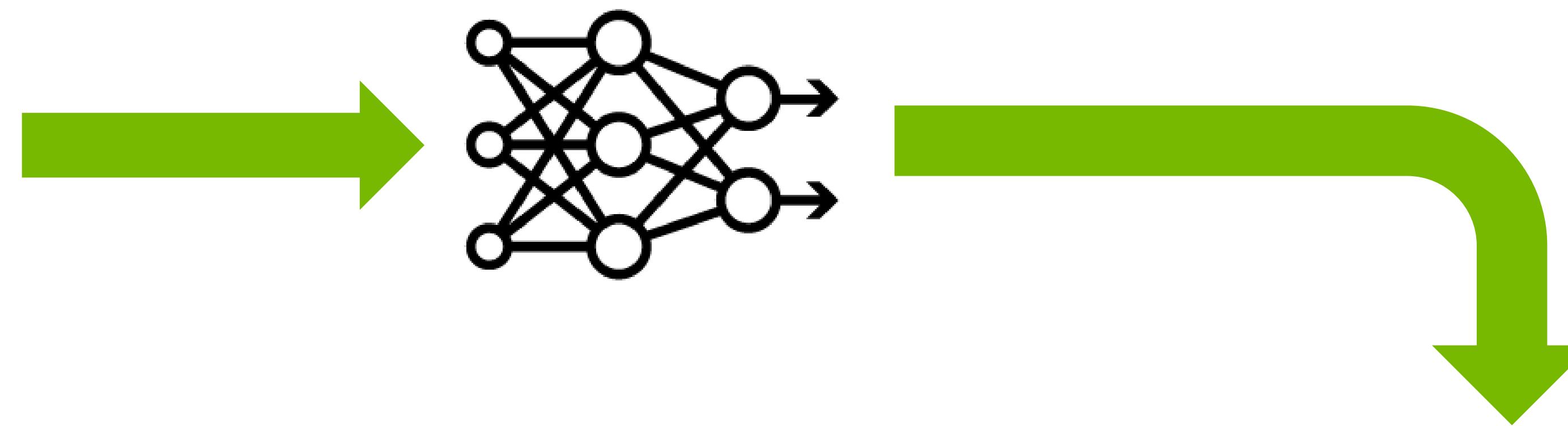
- Part 5: CLIP

- Part 6: Wrap-up & Assessment

Contrastive Language-Image Pre- Training (CLIP)

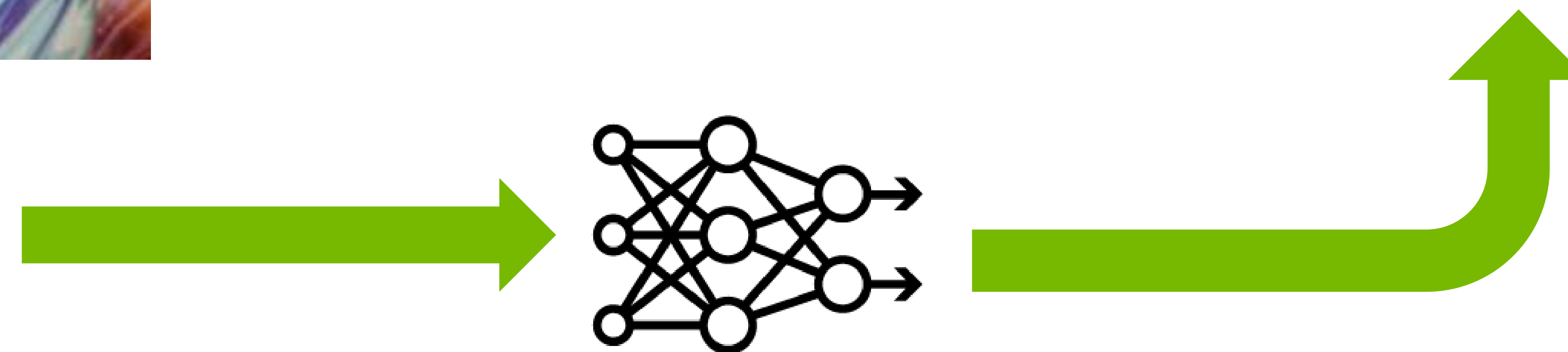
Matching Text to Image

Is it Possible?

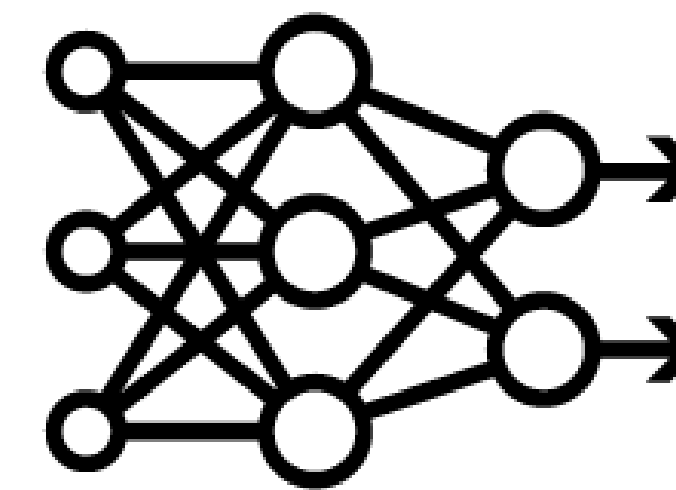


[0.8, -0.6, 0.7]

"A bunch of different marbles"

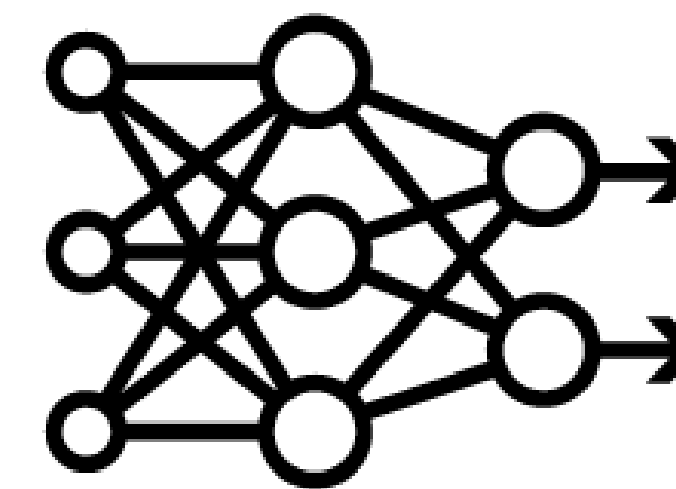


Cosine Similarity



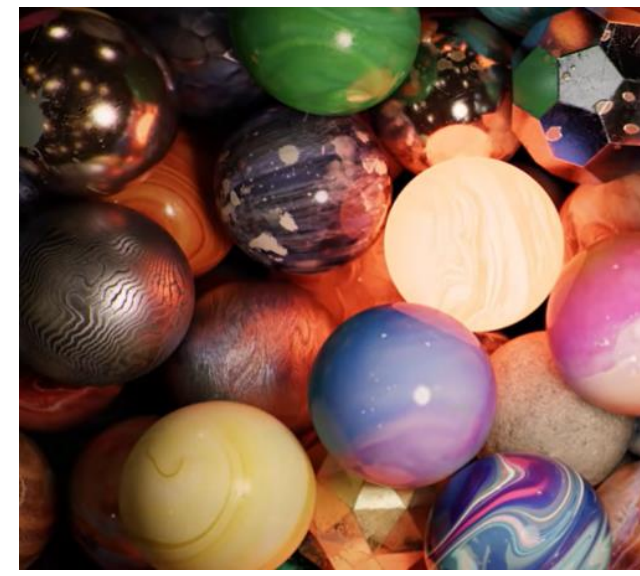
[0.0, 1.0]

"A bunch of different marbles"



[1.0, 1.0]

Cosine Similarity



[0.0, 1.0]

[1.0, 1.0]

"A bunch of different marbles"

45°

$$\cos(45^\circ) = \frac{\sqrt{2}}{2}$$

$$\cos(90^\circ) = 0$$

$$\cos(270^\circ) = 0$$

$$\cos(0^\circ) = 1$$

$$\cos(180^\circ) = -1$$

Dot Product



$[0.0, 1.0]$

$[1.0, 1.0]$

“A bunch of different marbles”

$[0.0, 1.0]$

$[1.0, 1.0]$

Dot Product



[0.0, 1.0]

[1.0, 1.0]

“A bunch of different marbles”

	A	B	A x B
x	0	$\frac{\sqrt{2}}{2}$	0
y	1	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$

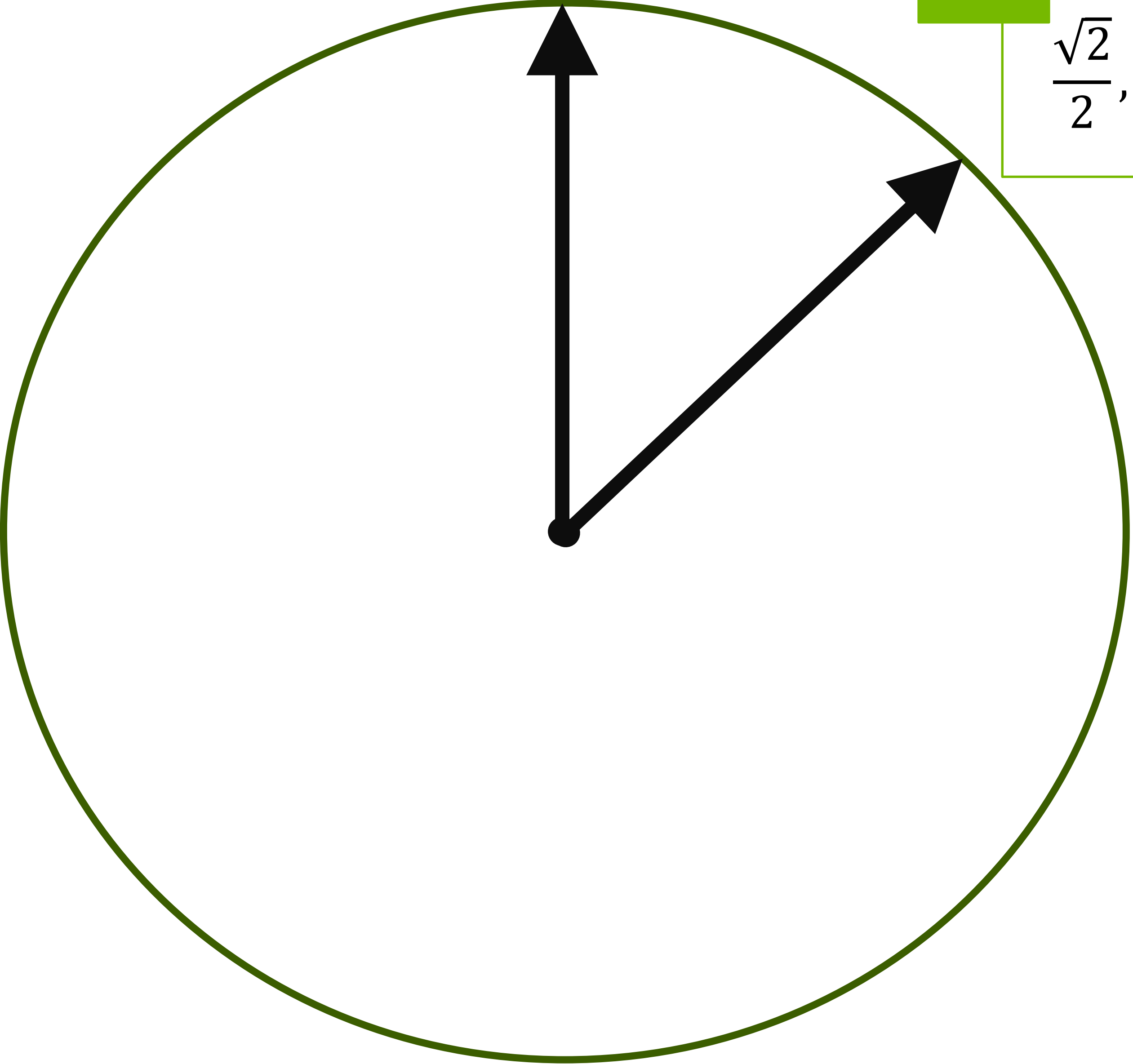
$\cos(45^\circ) = \frac{\sqrt{2}}{2}$

A

[0.0, 1.0]

B

$\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}$



CLIP Training

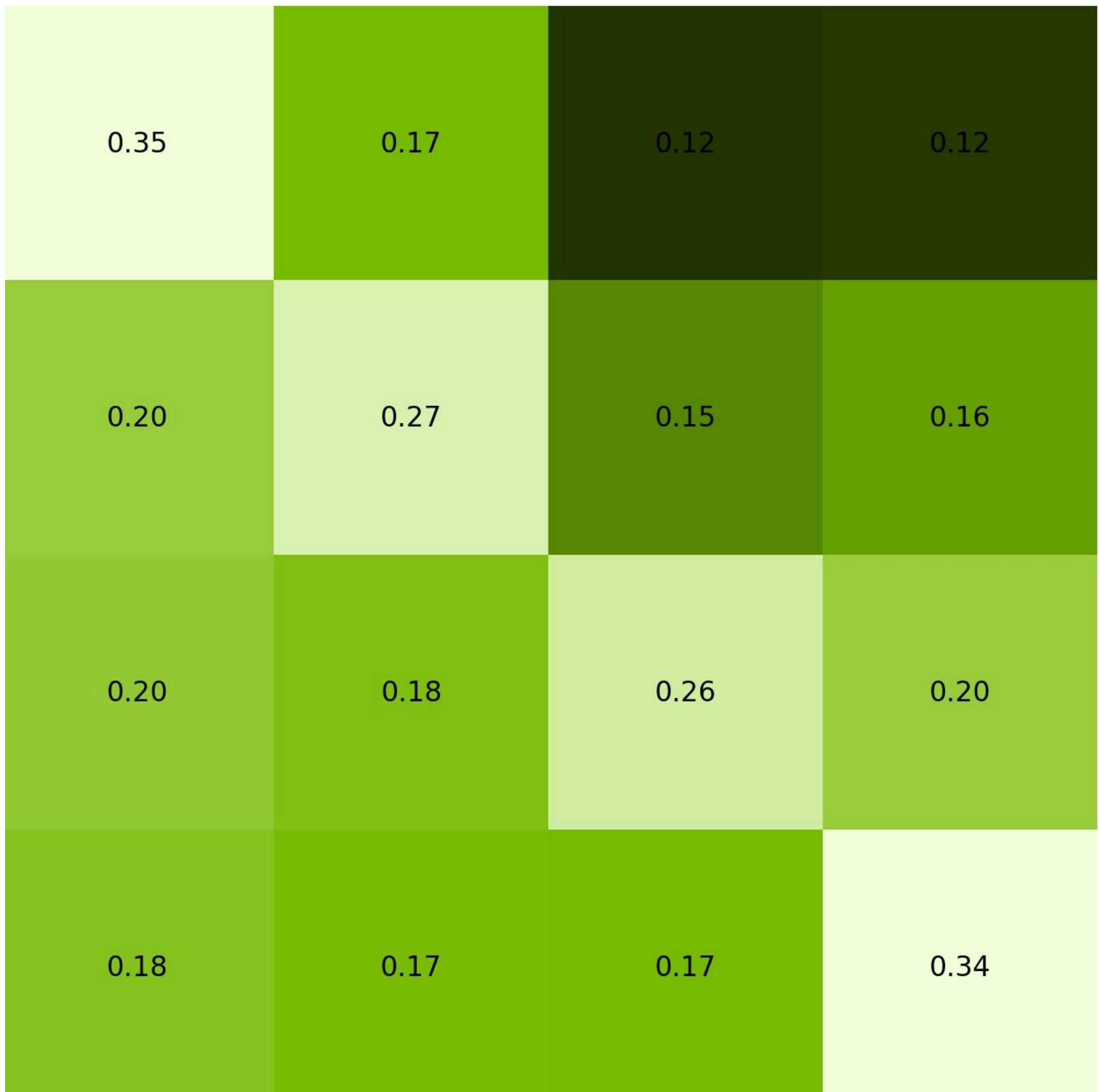


A happy corgi

A leather jacket

A green spiral

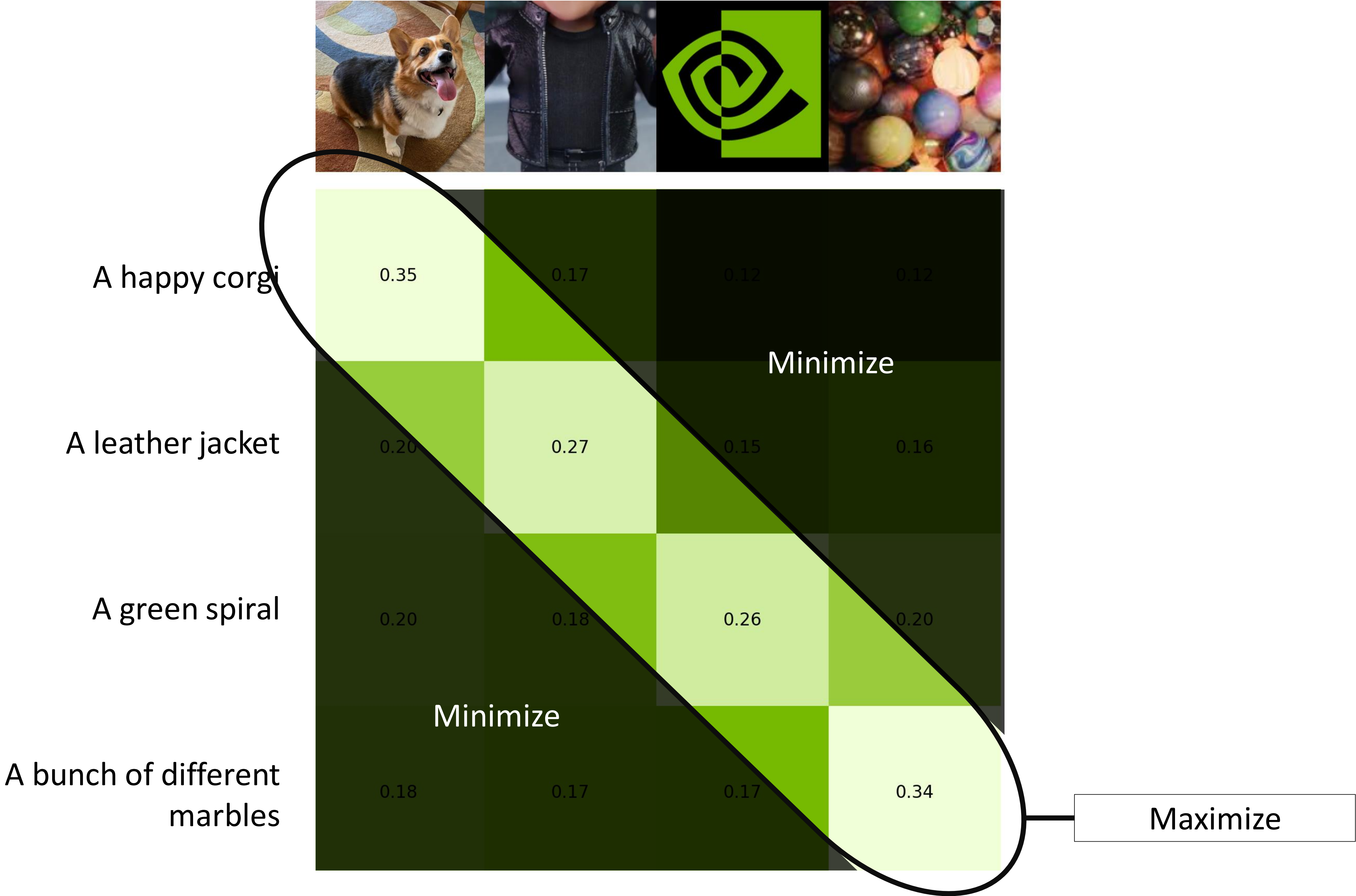
A bunch of different marbles



Cosine similarity between encoding for “A happy corgi” and encoding for each image

Cosine similarity between encoding for the NVIDIA logo and encoding for each text

CLIP Training





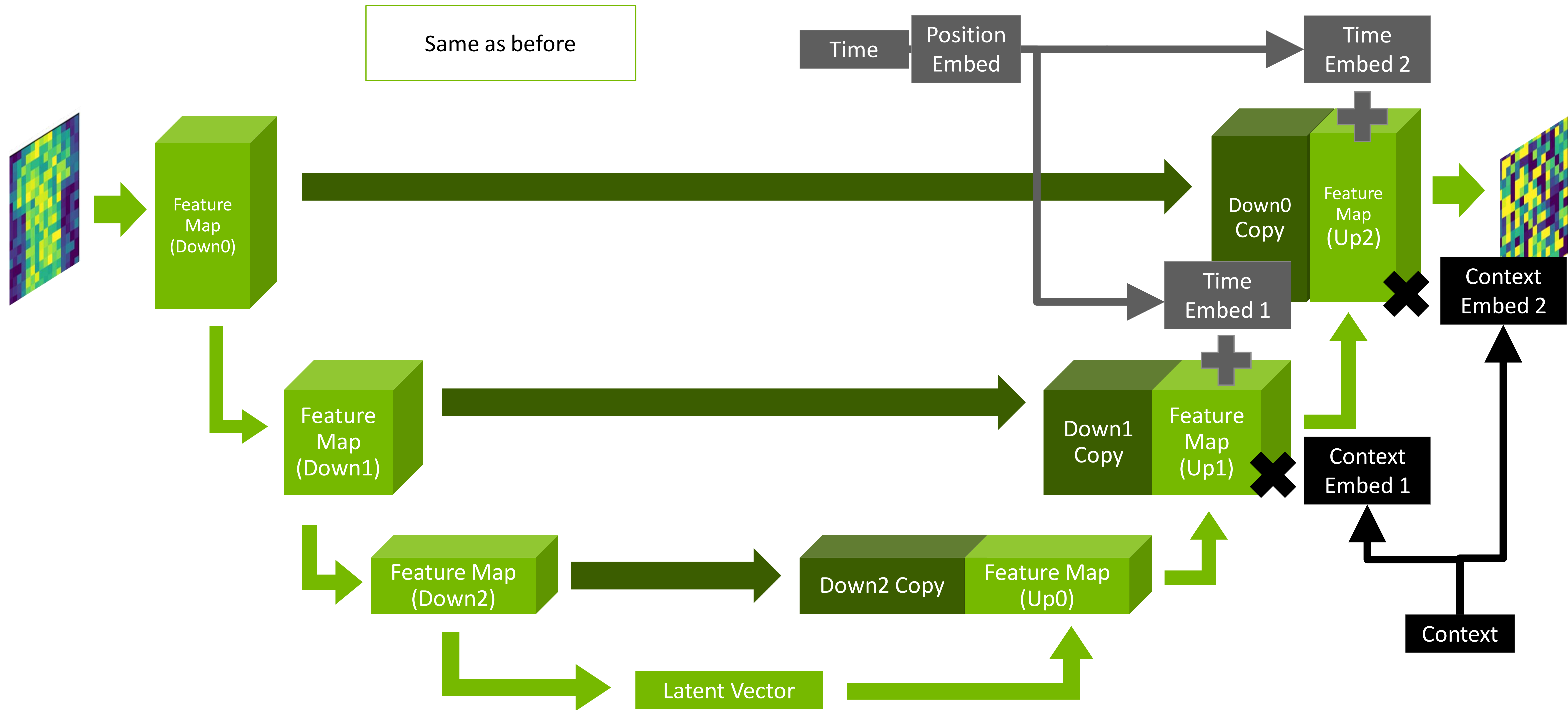
An Experiment

“

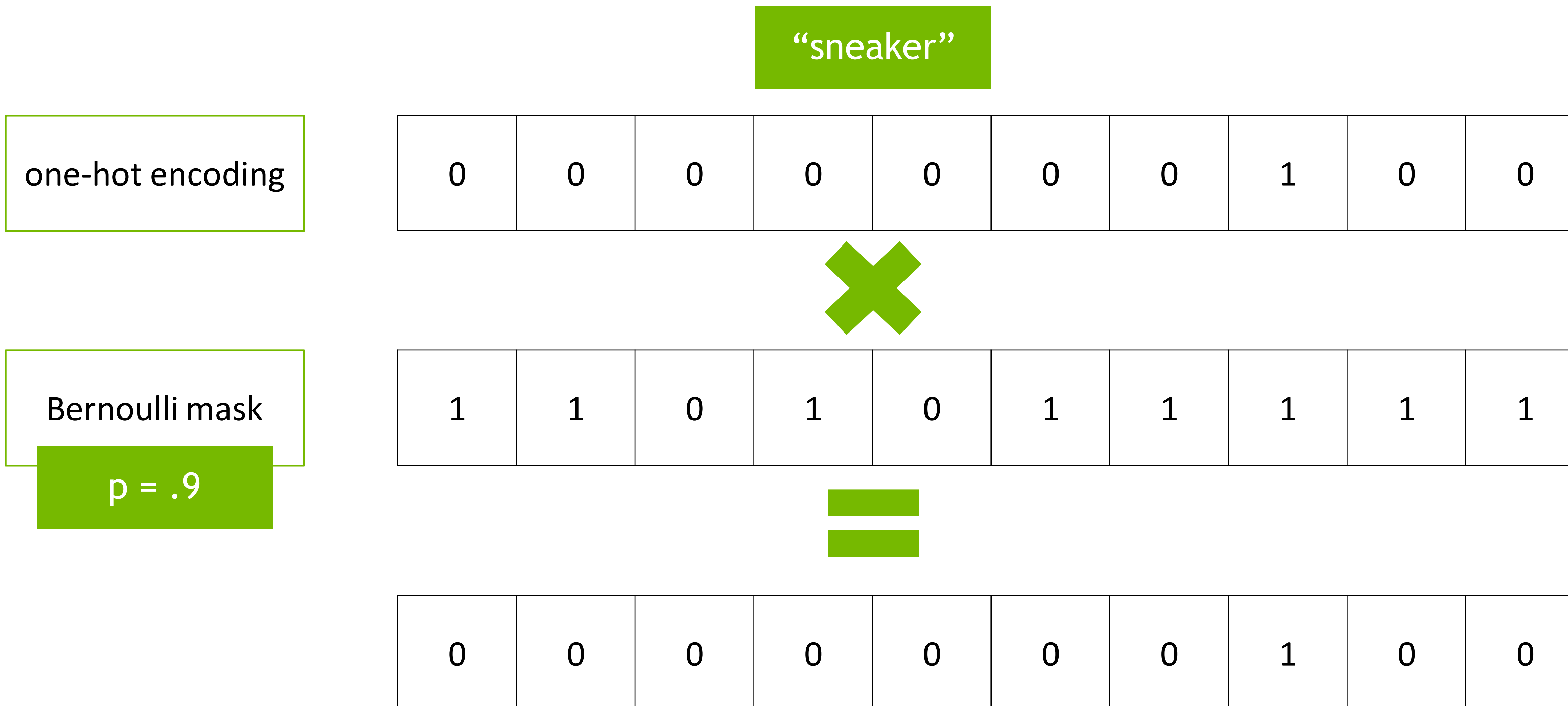
If CLIP is a pretrained model, do we need text labels to make a text-to-image model?

”

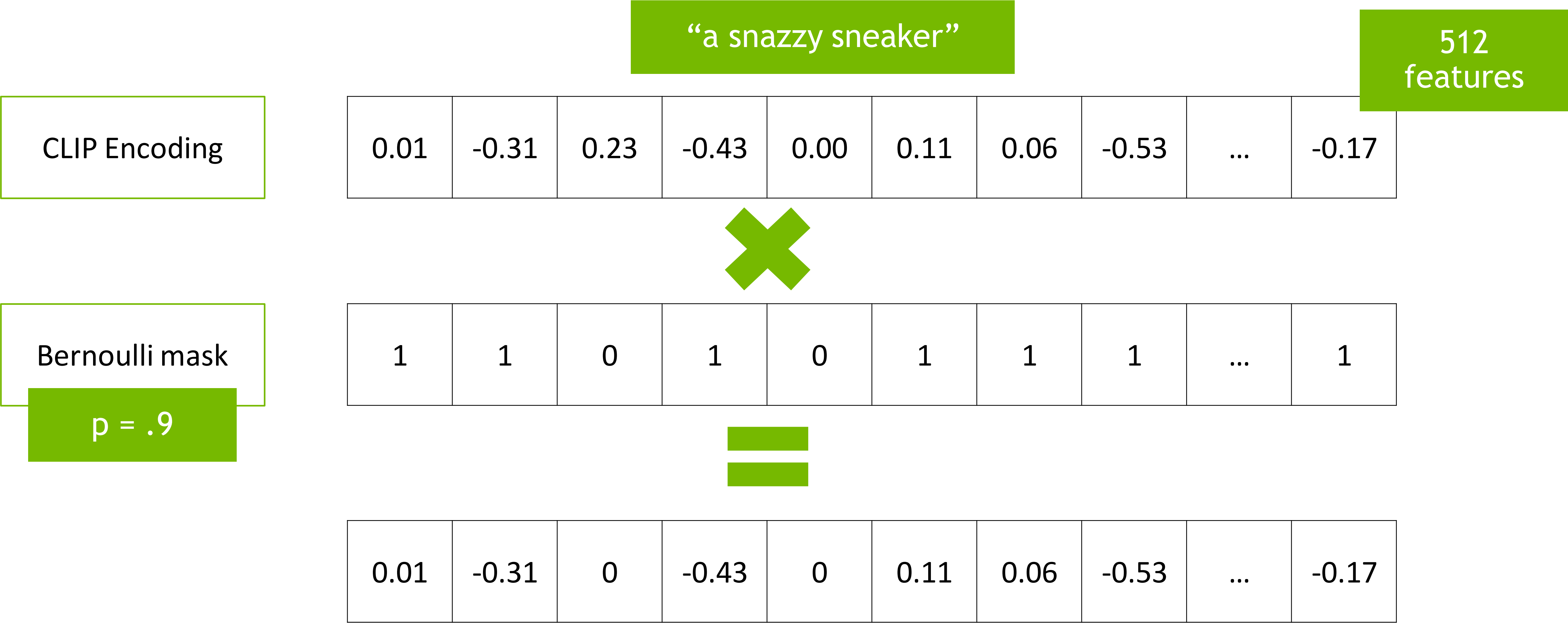
The Final Model



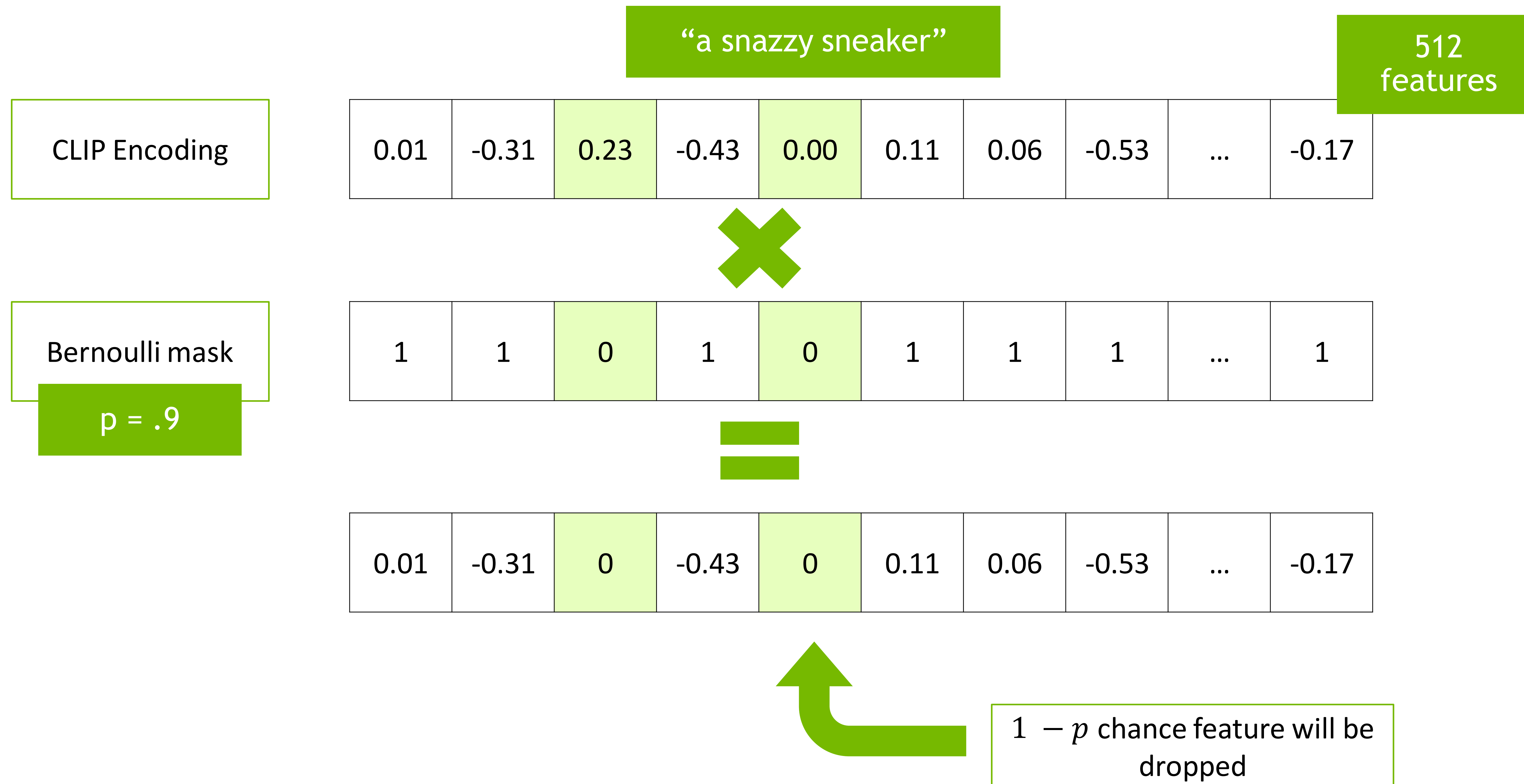
From Class to Context



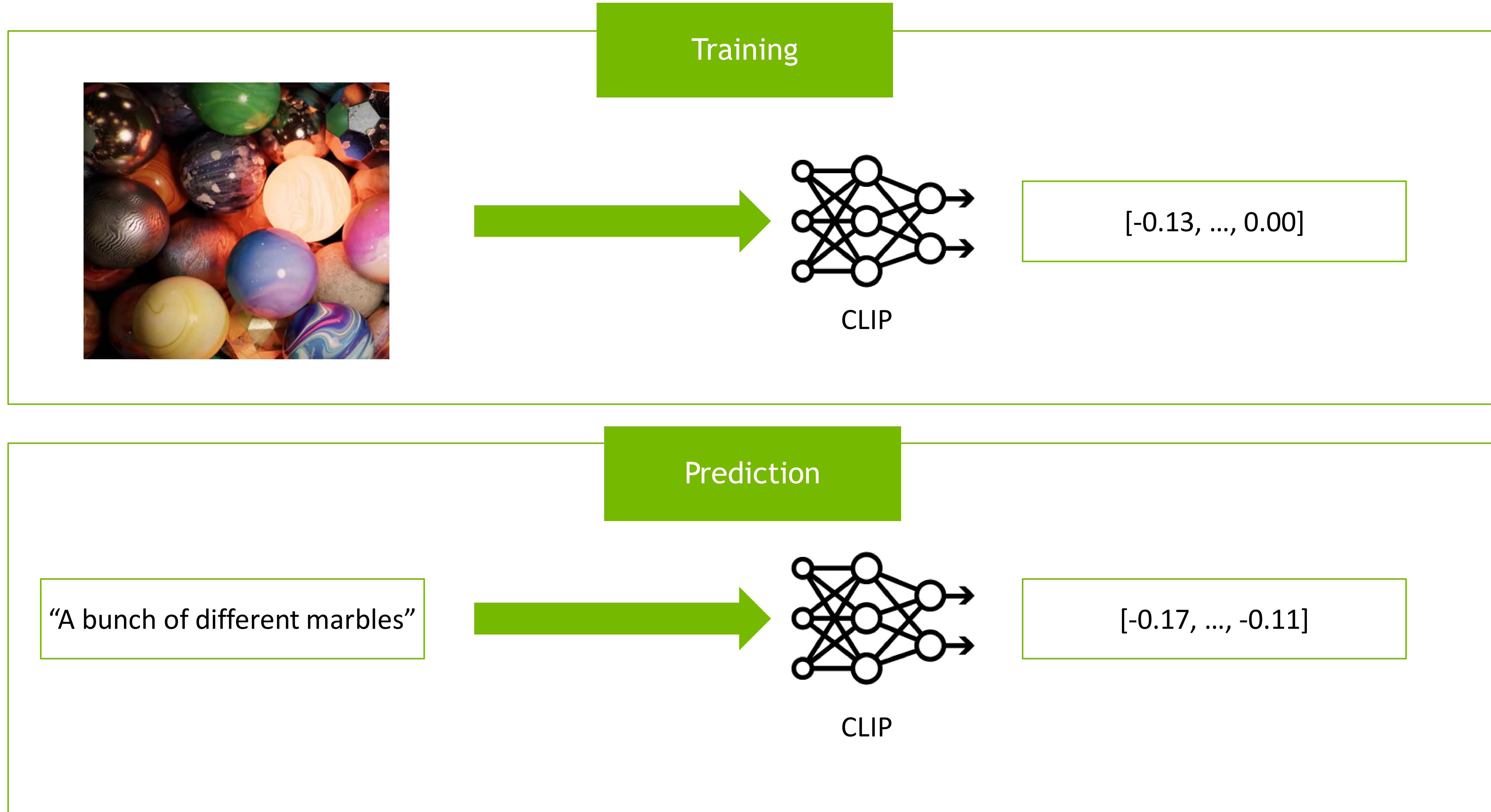
From Class to Context



From Class to Context



Experimenting with CLIP



The background features a series of parallel diagonal lines in various shades of green, creating a sense of depth and movement. Overlaid on these lines are several overlapping, rounded rectangular shapes in different green tones, some appearing to be layered on top of others. The overall effect is a modern, abstract design.

Let's get started!

