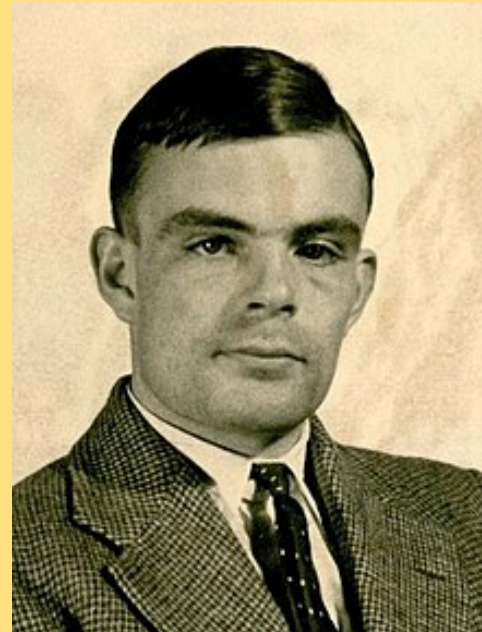


Asynchronous Perception Machines

Rajat, Yogesh,
BATMAN Hinton,
&

the warm-canadian-shadows.

LEGITIMATE



ALL CHARACTERS AND
EVENTS IN THIS SHOW--
EVEN THOSE BASED ON REAL
PEOPLE--ARE ENTIRELY FICTIONAL.

ALL CELEBRITY VOICES ARE
IMPERSONATED..... ~~BOOBY~~ **CHEAPLY**
THE
FOLLOWING PROGRAM CONTAINS
COARSE LANGUAGE AND DUE TO
ITS CONTENT IT SHOULD NOT BE
VIEWED BY ANYONE ■

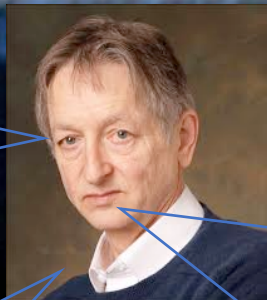
But first some **disclaimers**:

- This talk is **HIGHLY UN-PROFESSIONAL**
 - It contains **little-godzillas**.
 - And **the full force** of Star trek, star wars, too...
 - **Stonehenge** and **aliens** too.
 - It makes **jokes**.
 - It contains **weird roleplay** scenarios.
- **Last warning!!!**

Motivation

- A brief start with:

Work on GLOM
Kiddo.
Geoff



4 years AGO
Geoff sir,
Geoff sir,
I need Phd topic
Can't decide.

Next slide please



Hehe, it doesnt
work, but he does
not know it. 🍆

GLOoooooM
Sounds my cuppa
cake,
Whatcha GLOM?

- Jedi Huntron

- Jedi Warrior Huntron publishes paper
 - 2021



How to represent part-whole hierarchies in a neural network

Geoffrey Hinton

This paper **does not describe a working system**. Instead, it presents a single idea about representation which all

- We want to **make this work**.

- Why? Hinton is just a “Crazy Old Nut”.



- Not Gen Z.



- But, seriously: Why work on GLOM?

According to Hinton's long-time friend and collaborator Yoshua Bengio, a computer scientist at the University of Montreal, if GLOM manages to solve the engineering challenge of representing a parse tree in a neural net, it would be a feat—it would be important for making neural nets work properly. “Geoff has produced amazingly powerful intuitions many times in his career, many of which have proven right,” Bengio says. “Hence, I pay attention to them, especially when he feels as strongly about them as he does about GLOM.”

- That's why...



Geoffrey Hinton has a hunch about what's next for AI

A decade ago, the artificial-intelligence pioneer transformed the field with a major breakthrough. Now he's working on a new imaginary system named GLOM.

- So what, hunches have no REAL value.
- They are NOT publishable....
- Who CARES about arxiv. It's NOT peer-reviewed.

- NeurIPS'24.

OpenReview.net

← Back to Author Console

It now appears that some of the ideas in GLOM could be made to work.

<https://www.technologyreview.com/2021/04/16/1021871/geoffrey-hinton-glom-godfather-ai-neural-networks/>

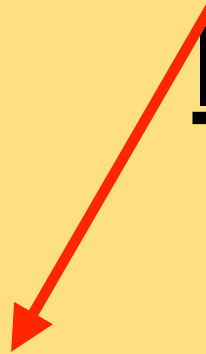
- So what, we already have SOTA?

The Forward-Forward Algorithm: Some Preliminary Investigations

Geoffrey Hinton
Google Brain
geoffhinton@google.com

¹¹There is clearly no problem adding skip-layer connections, but the simplest architecture is the easiest to understand and the aim of this paper is understanding, not performance on benchmarks.

And NOW, ladies and gentlemen,
It's GANGSTA time, 🗡️ 🦴



PRESENTING
JEDI HUNTRON

And HUNT he shall,
HUNT He Shall

And Then,
Huntron takes out,
A lightsaber

Wait.. 🙄

it's slightly discharged!!

Okay, Here we go:



geoffhinton OP · 10y ago · Google Brain

There has been recent ma
 "holes" you can create in
 not with the width of a lay
 of RBMs is quite closely r
 math is not my thing.



It includes contrastive self-supervised learning and performs hierarchical segmentation as a part of recognition rather than as a separate task. **No more** boxes.

A second advantage of GLOM is that it does not require dynamic **routing**. Instead of routing information from a part capsule to a specific capsule that

¹⁰This solves a version of Hilbert's 13th problem.

it is **no more** paradoxical than a surfer wh

recent tricks [Grill et al., 2020, Chen and He, 2020] it may **eliminate** the need for negative examples.

By allocating neurons to locations rather than to **GLOM eliminates** a major weakness of capsule models, the good aspects of these models:

propagation is required. One is that it **eliminates** the problems. It also eliminates mode collapse.

⁸Adam Kosoriek suggested using universal capsules in 2019, but I was put off by the **symmetry** breaking issue and failed to realise the importance of this approach.

This paper **does not** describe a working system.



- **No more** data-augmentation
- **No more** encoder-decoder
- **No more** pretext-task
- **No more** softmax
- **No more** parallel-perception
- **No more** routing
- **No more** boxes
- **No complex math.** Just backprop.

- Revive an OLD mechanism called folding-unfolding.

- What's Next:

1 What is wrong with backpropagation

...of a deep neural net to absorb a lot of information from a single training case by jumping to a set of weights that handles that case perfectly could be of interest to psychologists who are tired of creeping down gradients²⁰

Shorry, it sorta broke,



Some Demonstrations of the Effects of Structural Descriptions in Mental Imagery*

GEOFFREY HINTON
University of California, San Diego

A visual imagery task is presented which is beyond the limits of normal human ability, and some of the factors contributing to its difficulty are isolated by comparing the difficulty of related tasks. It is argued that complex objects are assigned hierarchical structural descriptions by being parsed into parts, each of which has its own local system of significant directions. Two quite different schemas for a wire-frame cube are used to illustrate this theory, and some striking perceptual differences to which they give rise are described. The difficulty of certain mental imagery tasks is shown to depend on which of the alternative structural descriptions of an object is used, and this is interpreted as evidence that structural descriptions are an important component of mental images. Finally, it is argued that analog transformations like mental folding involve changing the values of continuous variables in a structural description.

Sep 15, 2017 - Technology

Artificial intelligence pioneer says ~~we need to start over~~

GAME



Steve LeVine

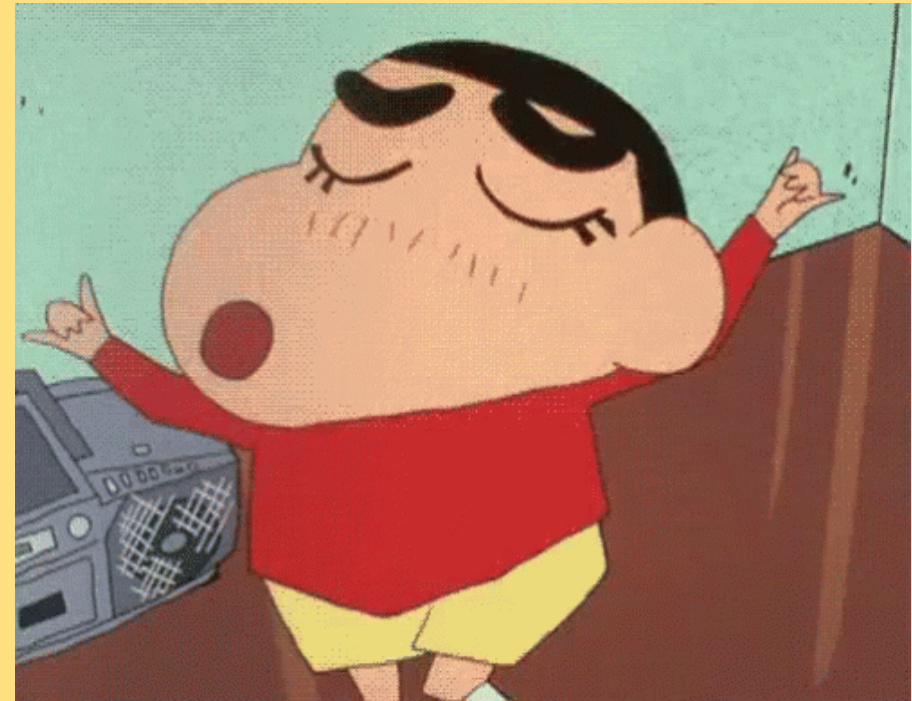
The bottom line: Other scientists at the conference said back-propagation still has a core role in AI's future. But Hinton said that, to push materially ahead, entirely new methods will probably have to be invented. "Max Planck said, 'Science progresses one funeral at a time.' The future



Anything else left

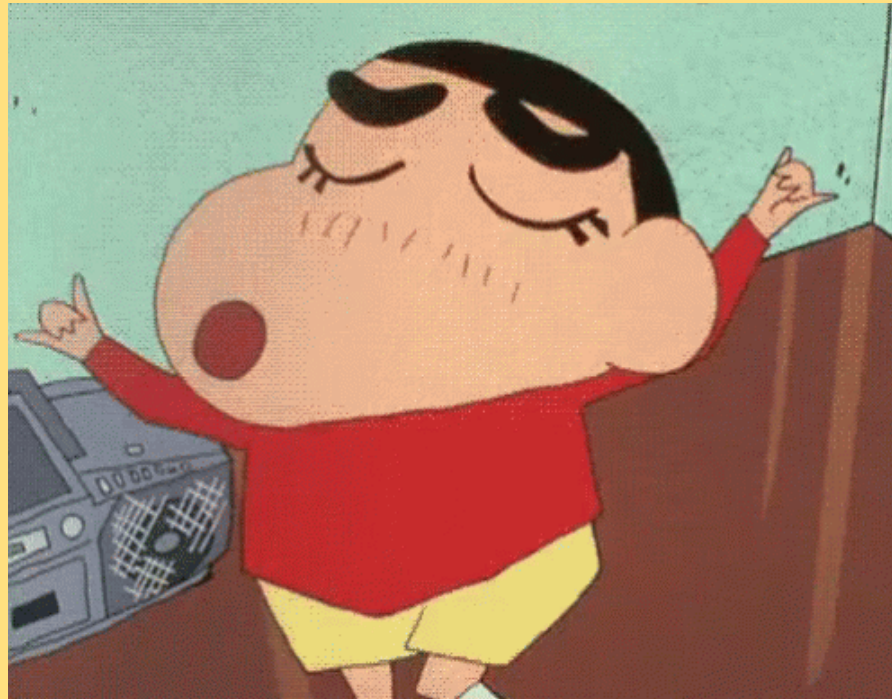
GEN Z?

I'm getting bored.



If, however, you are prepared to pay the energy costs required to run identical models on many copies of the same hardware, the ability to share weights across large models provides a much higher bandwidth way to share knowledge **than distillation** and may take intelligence to the next level.

- Maybe we can move **beyond DISTILLATION**
too



🎵 Old MacDonald
has a new farm, E-I-
E-I-O!



live long and prosper 🕊️

- What is **an issue** with Machine Perception?

- Scaling laws

- Take model
- Take a lot of data.
- **Learn good features.**
- Keep scaling up.
- Neural nets: Second law of thermodynamics > Laws of Linear Algebra.
- Accuracy pushes.
- Quantize the machine to 8 bits, roll out to real world.
- Amazing!!!! Isn't it.

Issues

- ROI seems to reduce i.e. increase in % of accuracy PER amount of parameter increase is reducing.
- **No way out of this scaling up problem.**
 - Problem: People **fighting** over getting cluster-time.
 - Training takes **forever**.
 - Sometimes **months**.
- We therefore need a **fundamental-fix**.

- ASSUMPTIONS
- Learning good features needs a lot of layers **stacked** over one other.

- The way out: Mortal Computation

Mortal Komputation: On Hinton's argument for superhuman AI.

I say it passes my bar for an interesting narrative. However, as a narrative, I don't consider it much stronger

- We want to bypass this **entirely**.
- Something which can run in a toaster. **Less than a dollar.**
- We can start calling them :
 - **Asynchronous Perception Machines**
 - **They have started working now.**
 - **Still a long way to go.**

- BREAKING ASSUMPTIONS

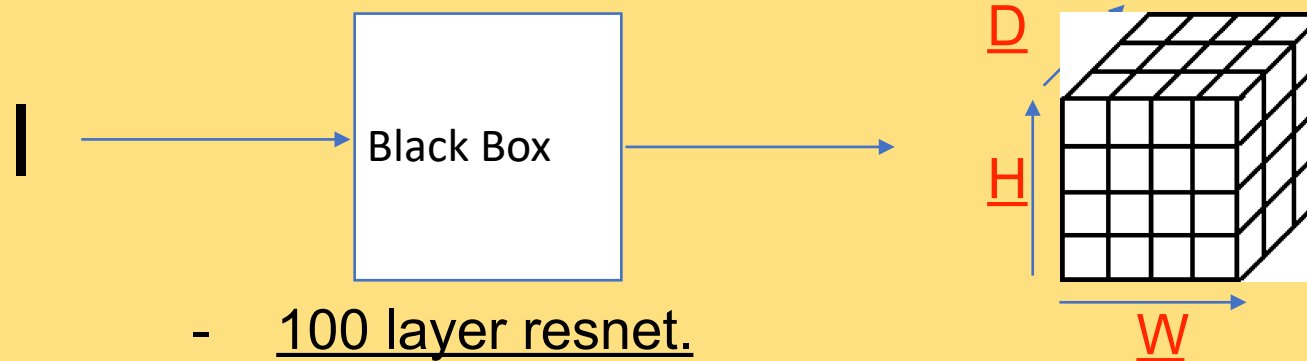
- Learning **good features** needs a lot of layers **stacked** over one other.

Where do features come from?

Geoffrey Hinton

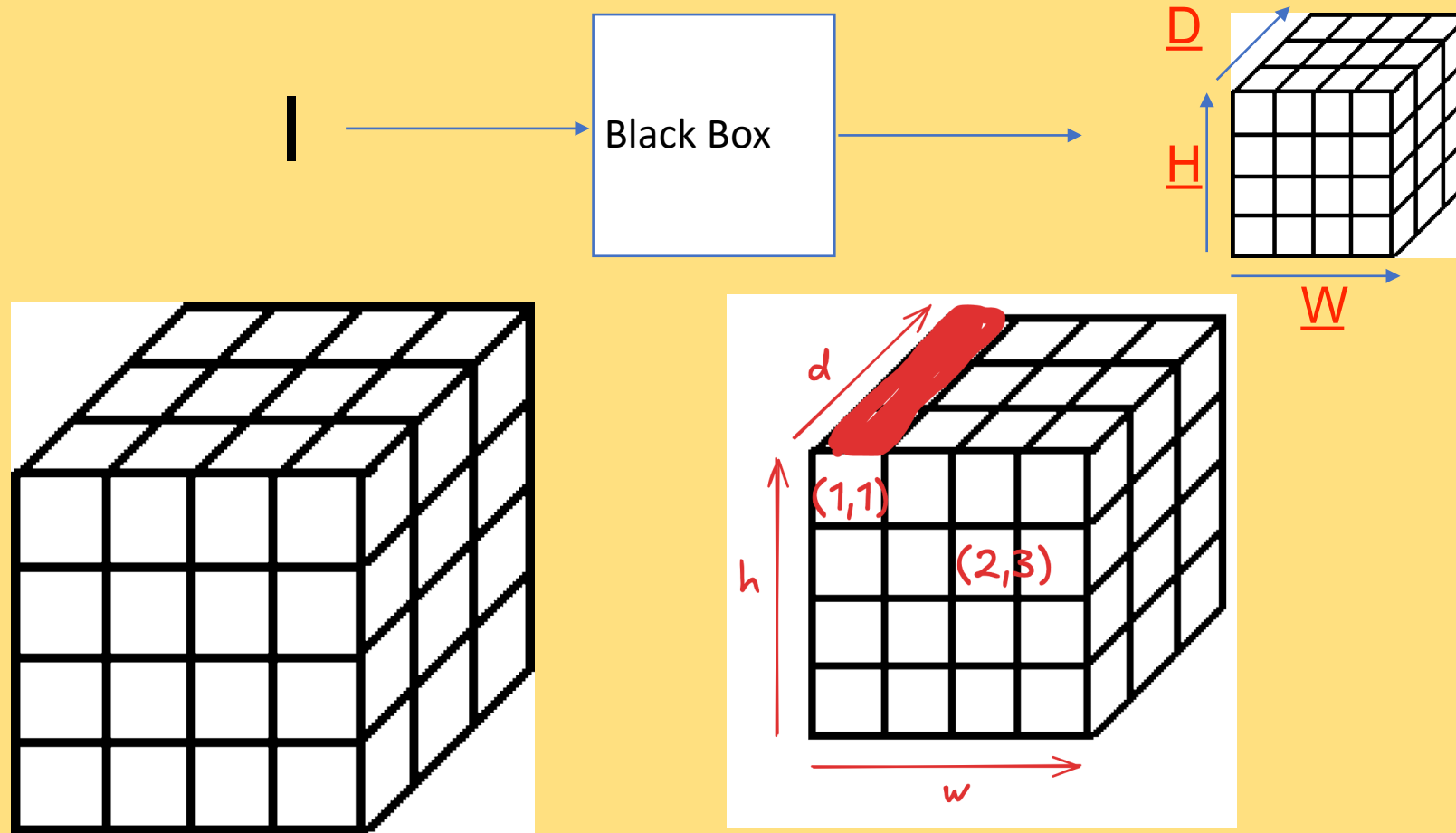
- **DUNNO** 😊
 - Let's assume whole thing is a **black-box**.

performed in the forward pass¹ in order to compute the correct derivatives². If we insert a black box into the forward pass, it is no longer possible to perform backpropagation unless we learn a differentiable model of the black box. As we shall see, the black box does not change the learning



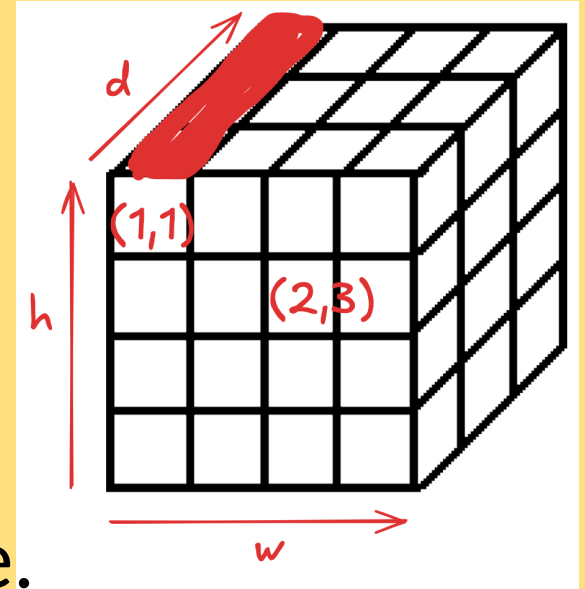
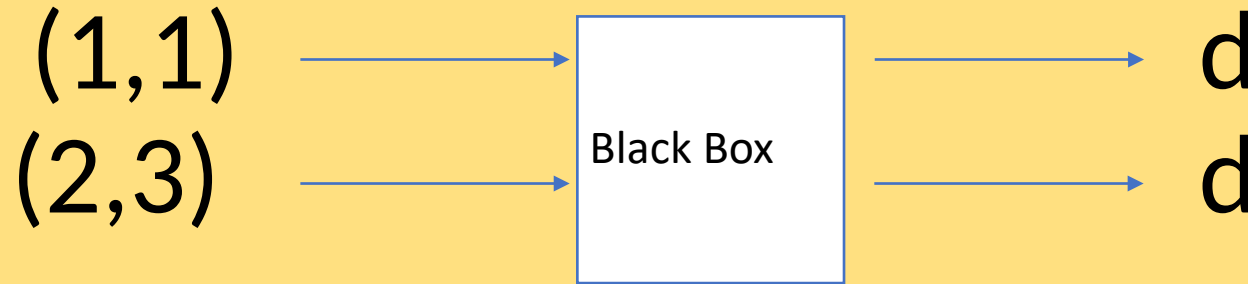
- 100 layer resnet.
- 12 layer ViT/transformer.
- Or a 1000 layer tiramisu :-)

- A reinterpretation of Feature Grid.



- Start thinking of this grid as d dimensional vector at each location.
- So there are $h*w$ such vectors.

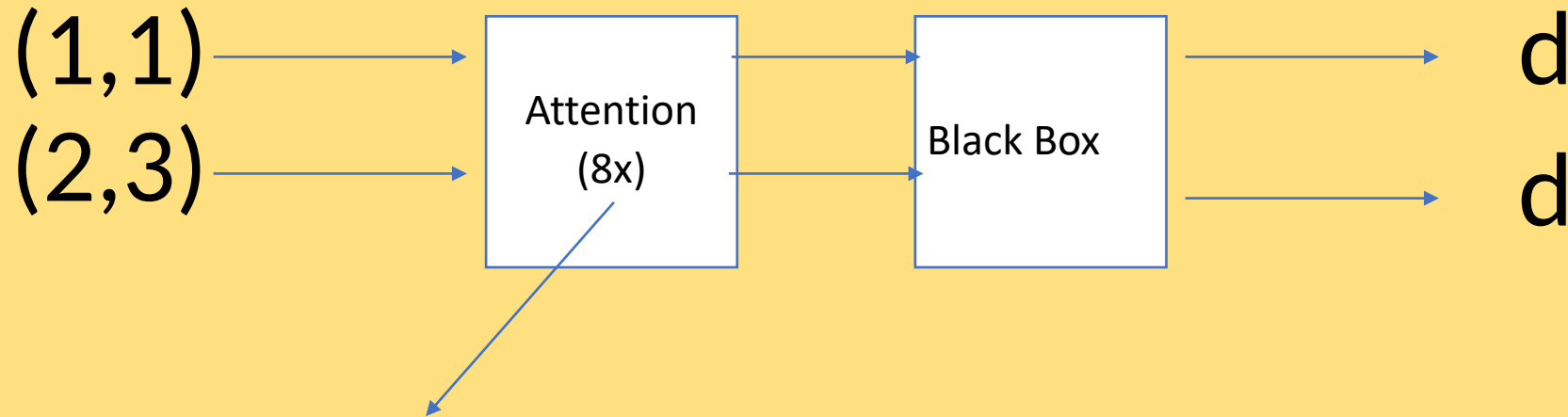
- So we can start imagining a new network.



- so you can query it $h * w$ times.
- you get a d dimensional number everytime.
- problem
- queries $(1,1)$, $(2,3)$ are independent.
- So since patches no longer communicate,
- there is no more possible way to machine perception.
- $H * w$ queries will make it slow.
- But it will be memory efficient

- Patches can no longer communicate

- “Classical Fix”

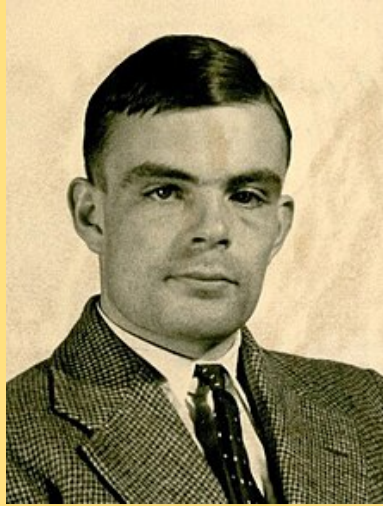


- Attention consumes memory.
- CNN is fine, but loses global-context since only runs on a window.
- We neither want a CNN, neither a transformer.
- Something new.
 - And we don't want patches to communicate among themselves.
 - That consumes too much memory!!!
- But we can't do machine perception without making patches communicate.
- See the paradox!!!
Impossible to get out of this ehhehh .

- And that was the assumption of Turing/GLOM



- **Attention:**
 - each eye is an attention head.
 - each head looks at all the tokens.
 - that consumes memory.



- **Turing:**

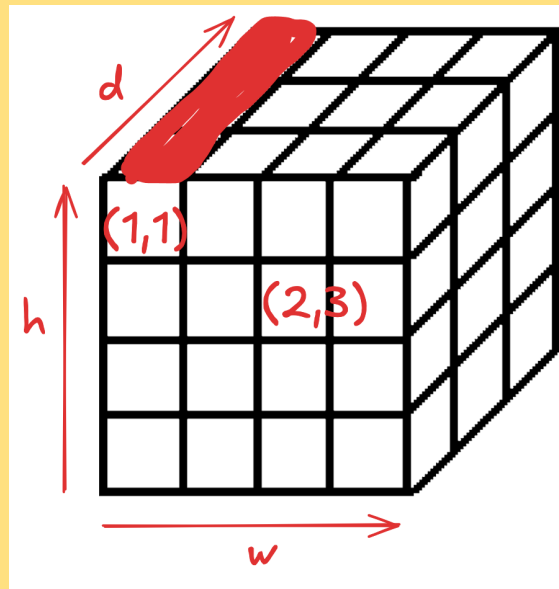
- different cells in the body/patches-in image communicate via blood, or substances.



- **GLOM:**

- make patches communicate to learn “islands of agreement”

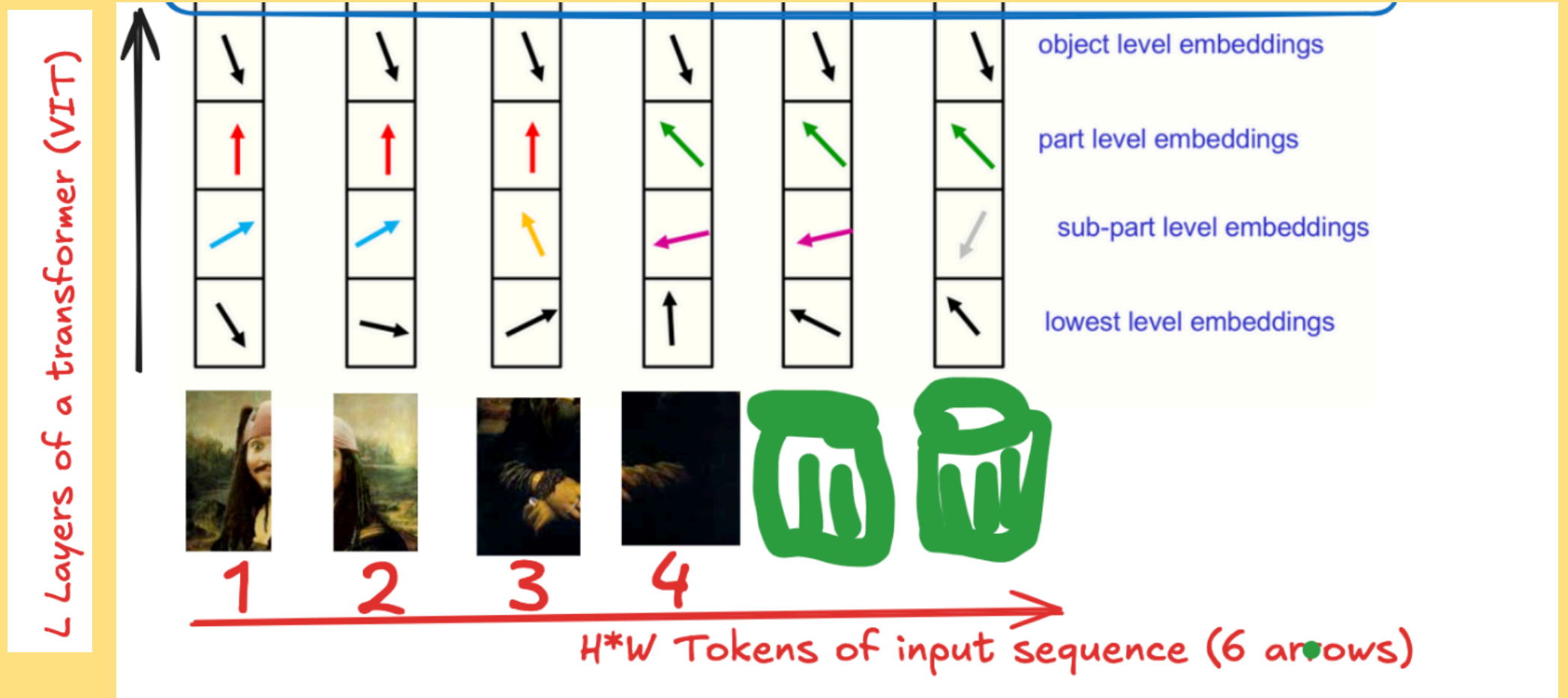
- And **NOW**, we will need another concept.
- We turn to **Jedi-Huntron**.



- **THINK OF THIS GRID AS VECTORS!!!!**

- [NeurIPS 2023] Hinton's Islands of Agreement

- So start thinking of features as little vectors/needles at each location.



- The only problem was that these islands of agreement were HYPOTHETICAL.

- And so we steal another idea:

The embedding vectors for nearby columns at a single time-step as GLOM settles

scene level embeddings
object level embeddings
part level embeddings
sub-part level embedding
lowest level embeddings

At each level there are islands of agreement. These islands represent the parse tree between levels, which makes it much more complicated.

It is a multi-level, real-valued Ising model with coordinate transforms between levels.

Stanford

Stanford CS25: V2 | Represent part-whole hierarchies in a neural network, Geoff Hinton

Stanford Online 672K subscribers

Subscribe

175

Share

Download

Save

The key to overcoming this apparent limitation of FF is to treat a static image as a rather boring video that is processed by a multi-layer recurrent neural network (Hinton, 2021). FF runs forwards in

- Algorithm for Hinton's islands of agreement:

The key to overcoming this apparent limitation of FF is to treat a static image as a rather boring video that is processed by a multi-layer recurrent neural network (Hinton, 2021). FF runs forwards in

- Take a **static image**.
- **Repeat** it many times.
- It becomes a **boring video**.
- Give it to a video transformer.
- Look at its third or fourth layer
- You will have a tensor of (H,W,D)
- Do t-sne on that, (H,W,3)
- And then visualize it.
- Video transformer is important. **We used Mvitv2.**

- [NeurIPS23] Hinton's Islands of Agreement.



- **No more** boxes. **No more** semantic supervision.
No more parametric upsamplers.

- Correction: Alan turing!!!!!!! 😊😊😊😊

convenient because it gives every cell its own private access to whatever DNA it might choose to express. Each cell has an expression intensity⁹ for each gene and the vector of expression intensities is similar for cells that form part of the same organ.

Suppose we want to predict

Nose



1



2



3



4



Suppose we want to predict

Nose

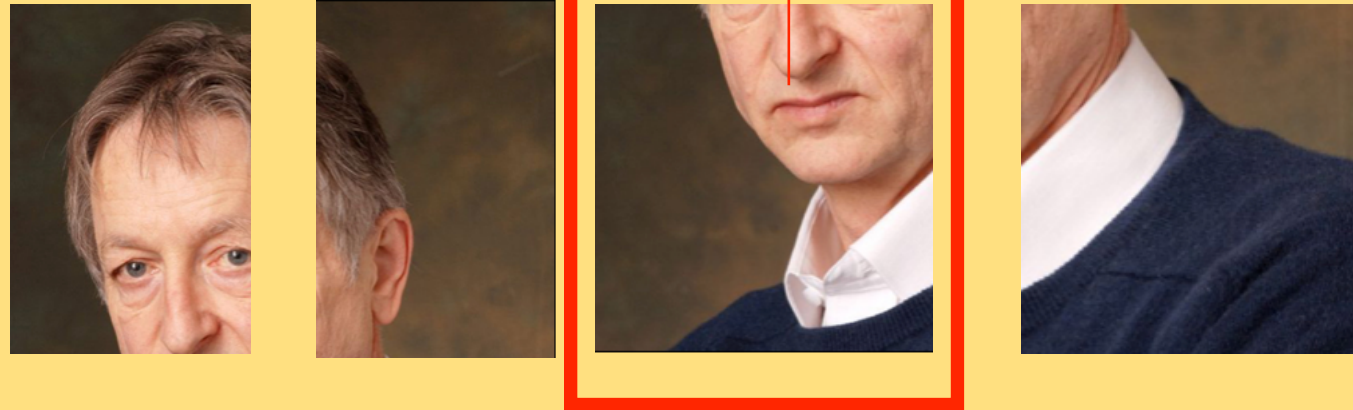


because it gives every cell its own private access to whatever DNA
choose to express. Each cell has an expression intensity⁹ for each gene

READ THIS AGAIN. READY???

Suppose we want to predict

Nose

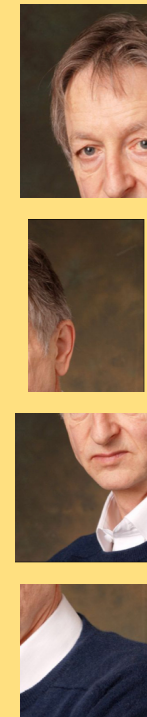
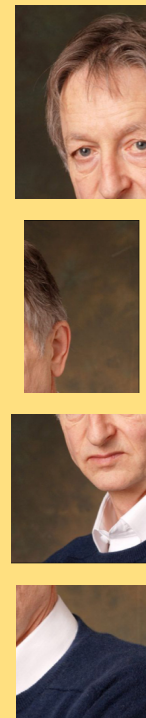
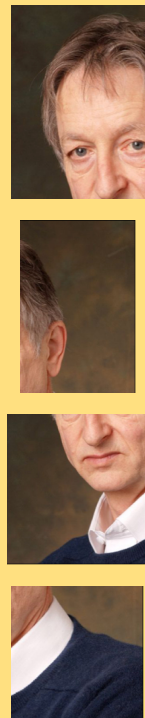
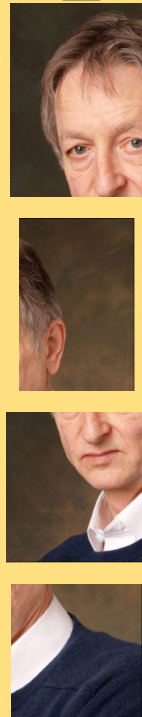


1

2

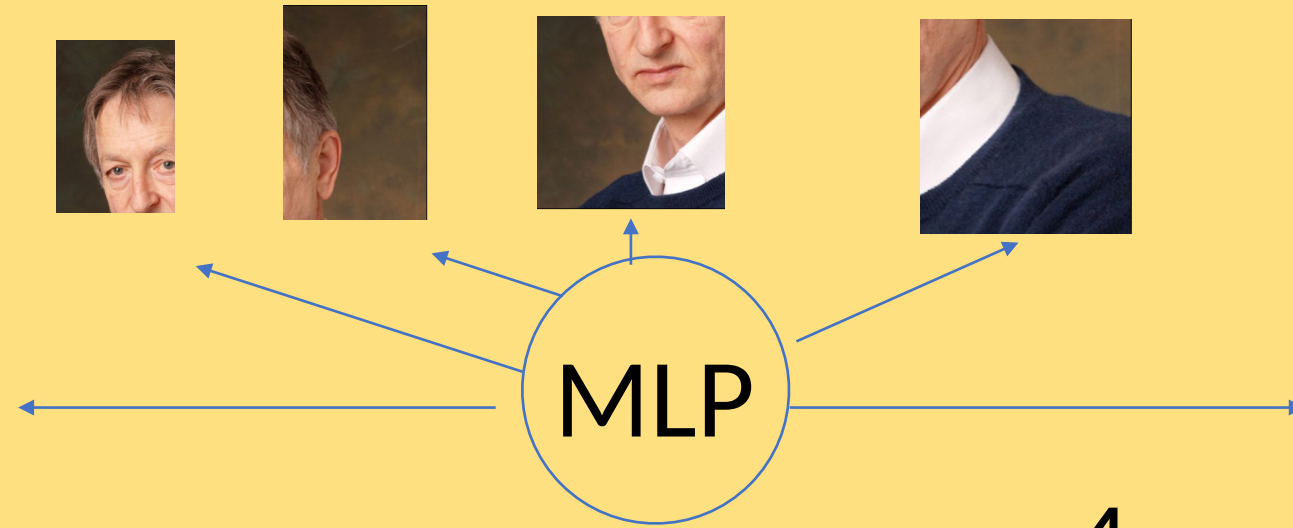
3

4



DNA =
HINTON's
Image

And the architecture becomes:

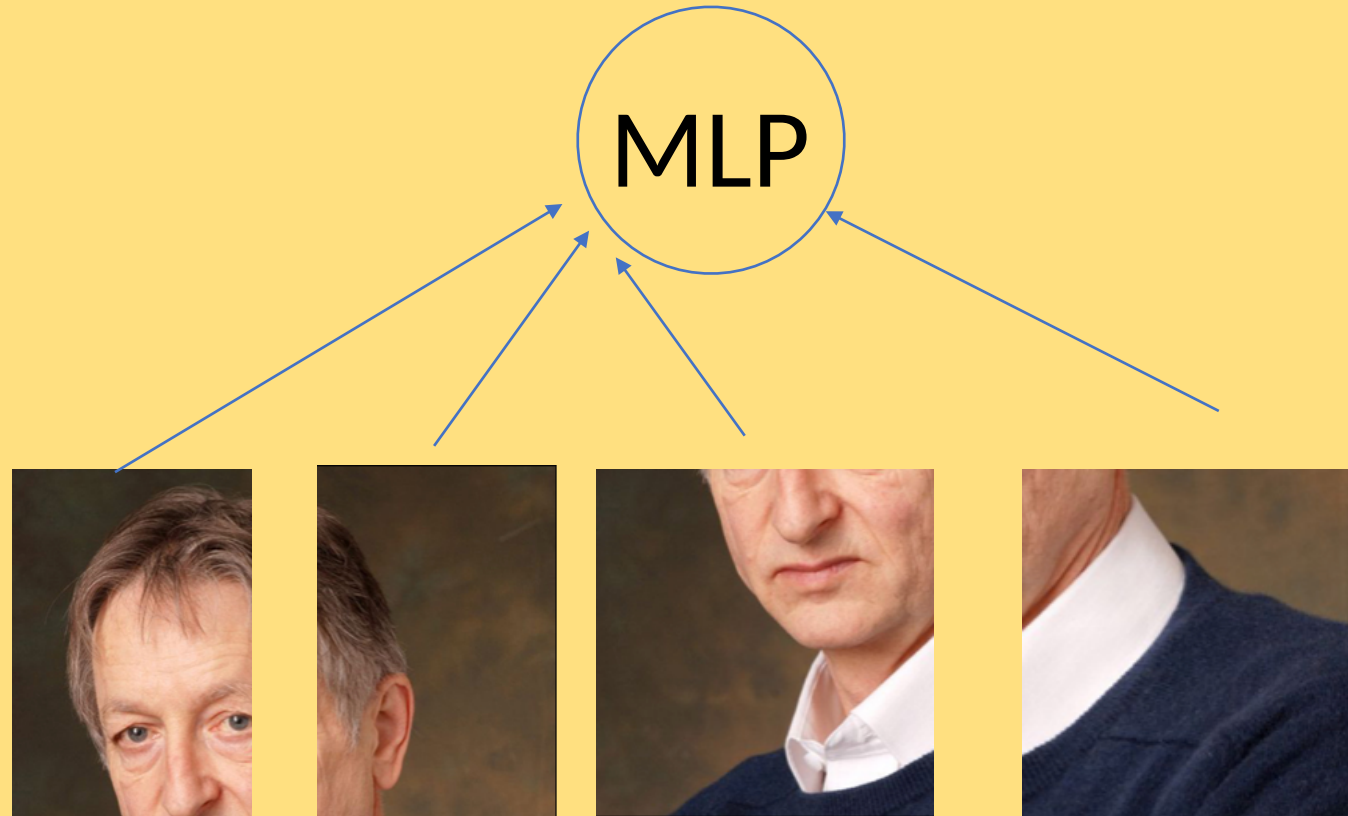


DNA =
HINTON's
Image

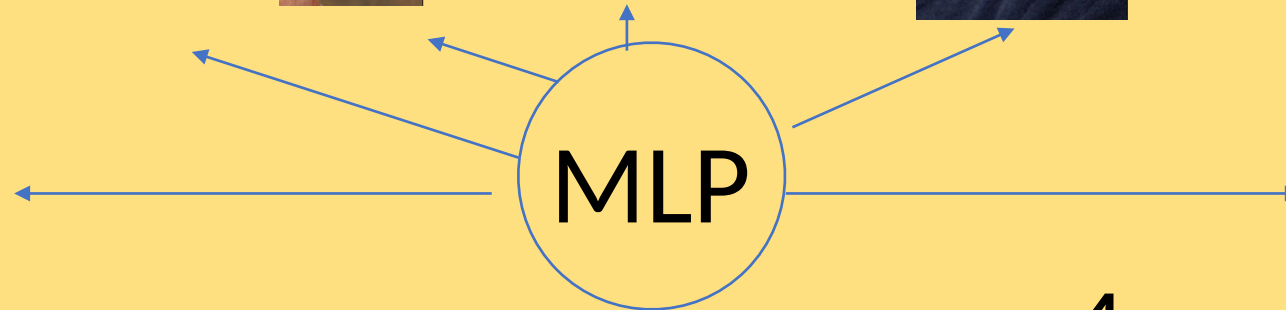
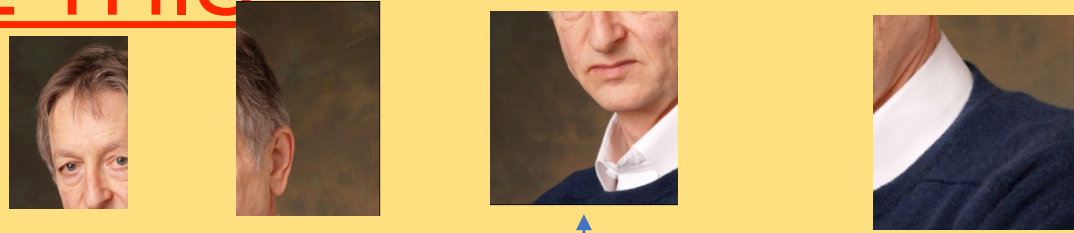


- THAT IS WHAT WE WERE SORTA FIXED.

- PPL WERE FEEDING IT LIKE THIS.



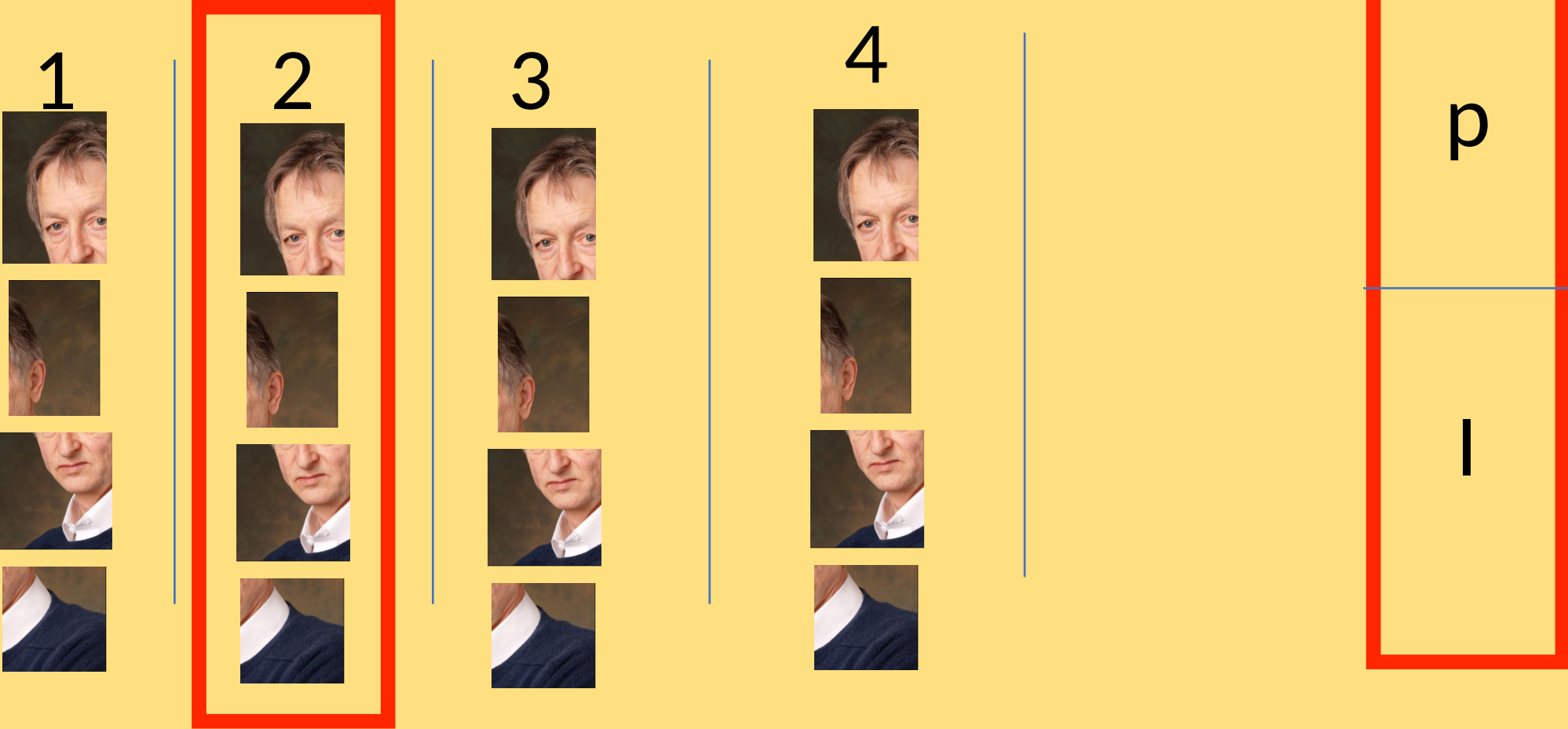
WE FED IT LIKE THIS



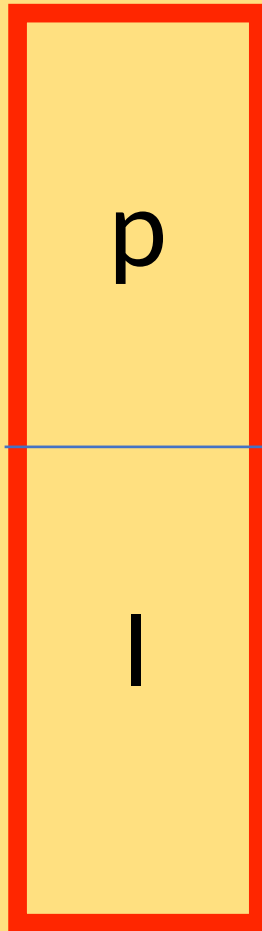
Now
look at
any
column :



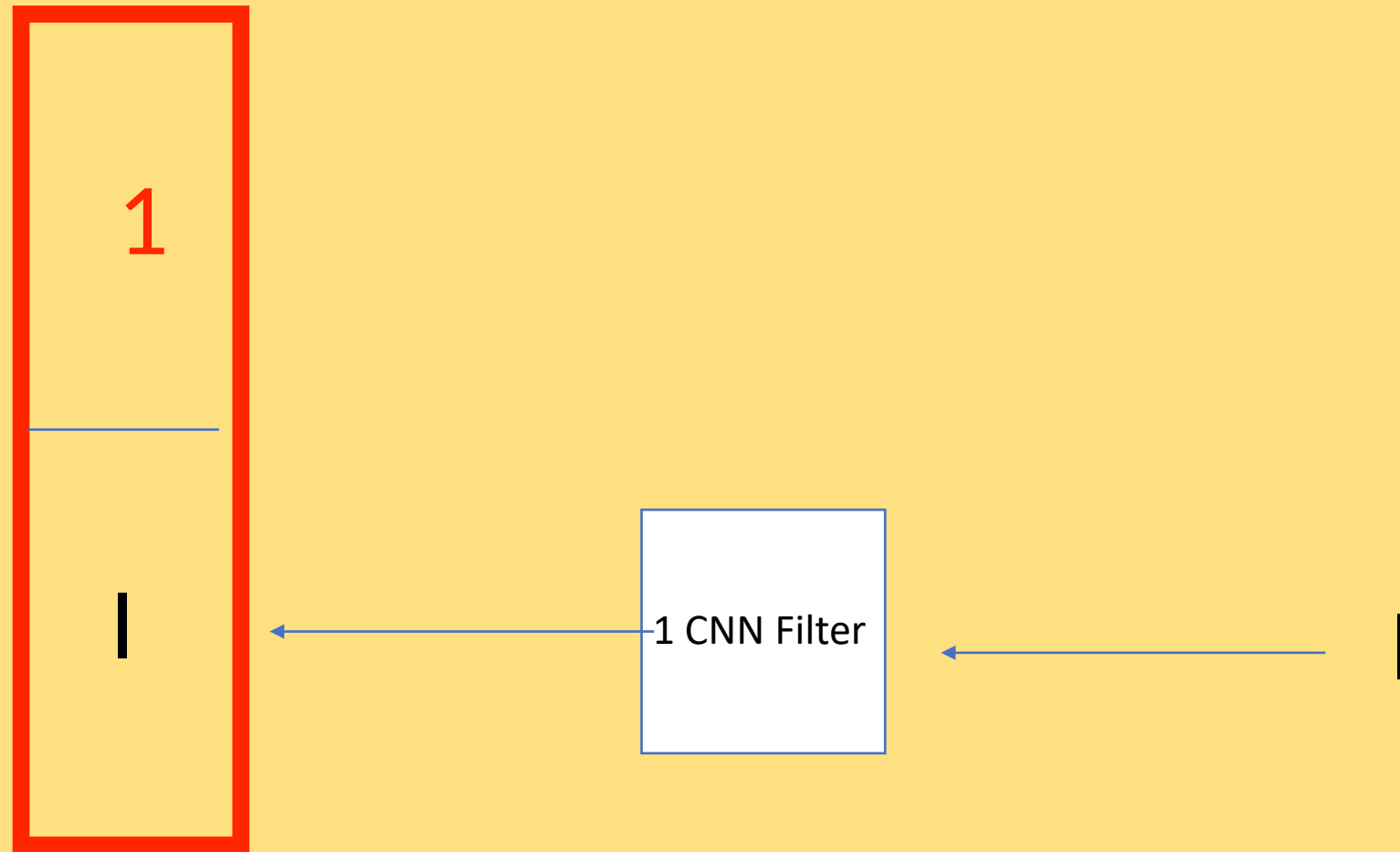
Here we will lay it out again:



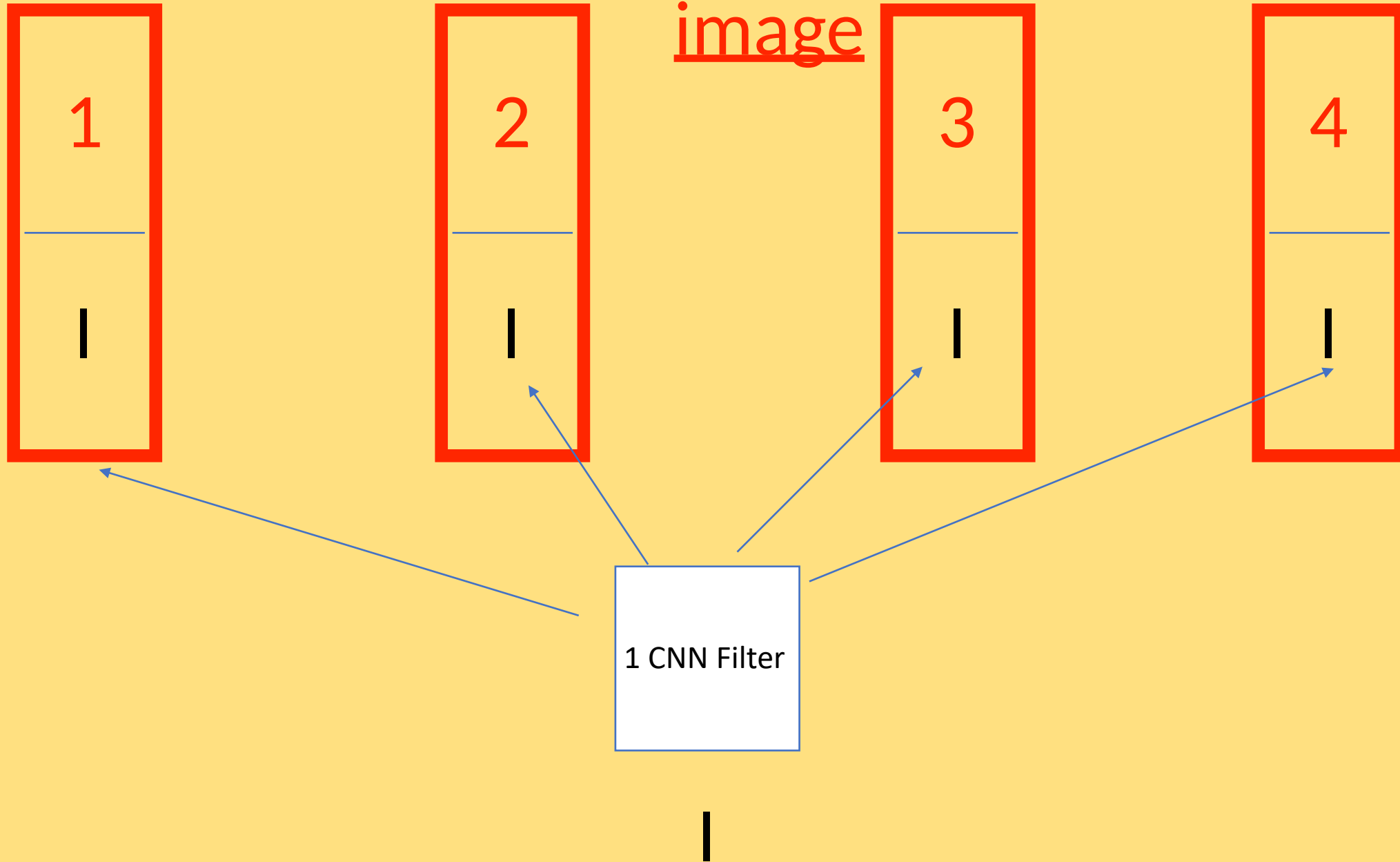
What we call as Trigger Column.



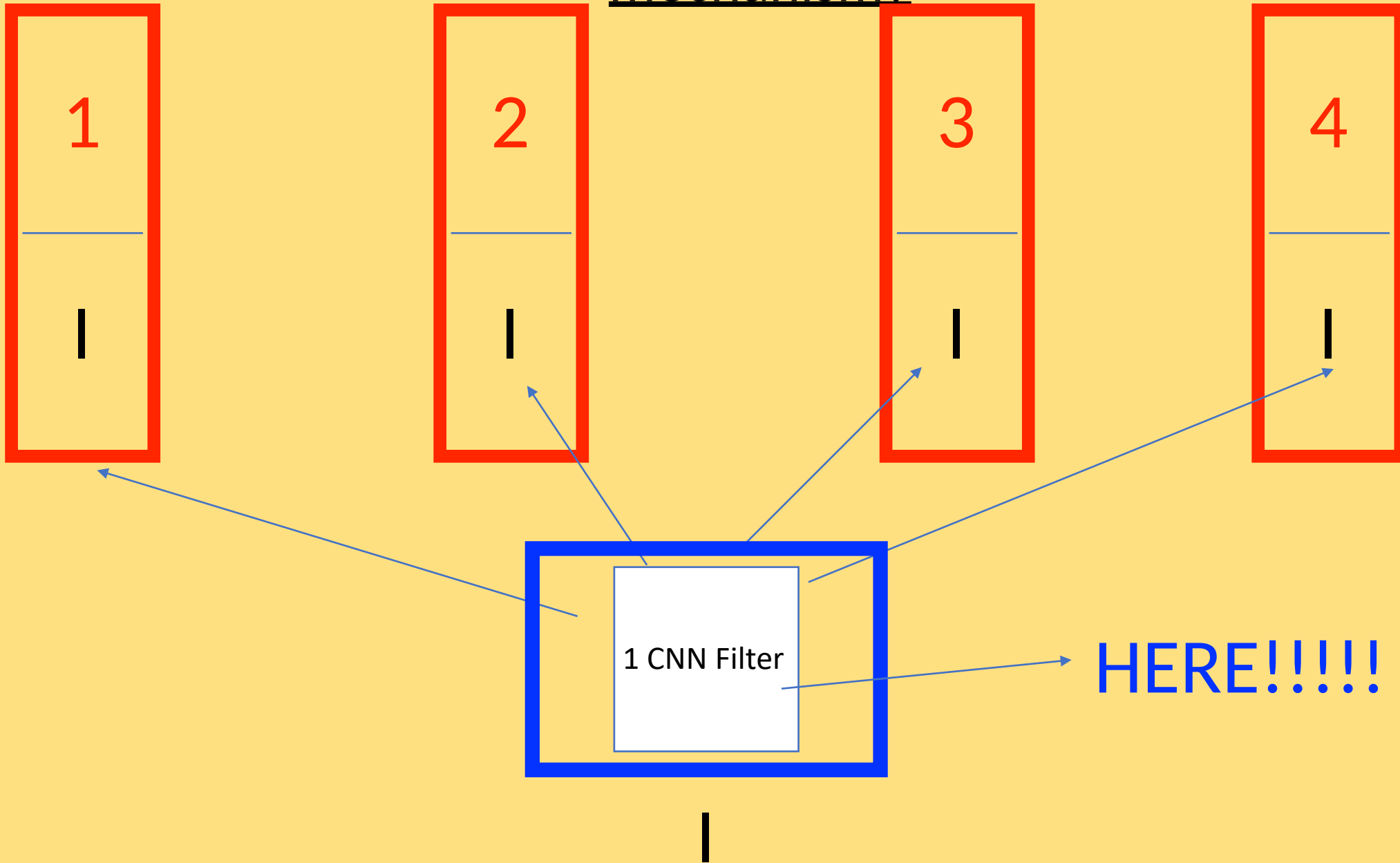
How does it work: The Unfolding



But we need four columns to **decode**



A q for you: Where are **the learnable parameters** in this mechanism?

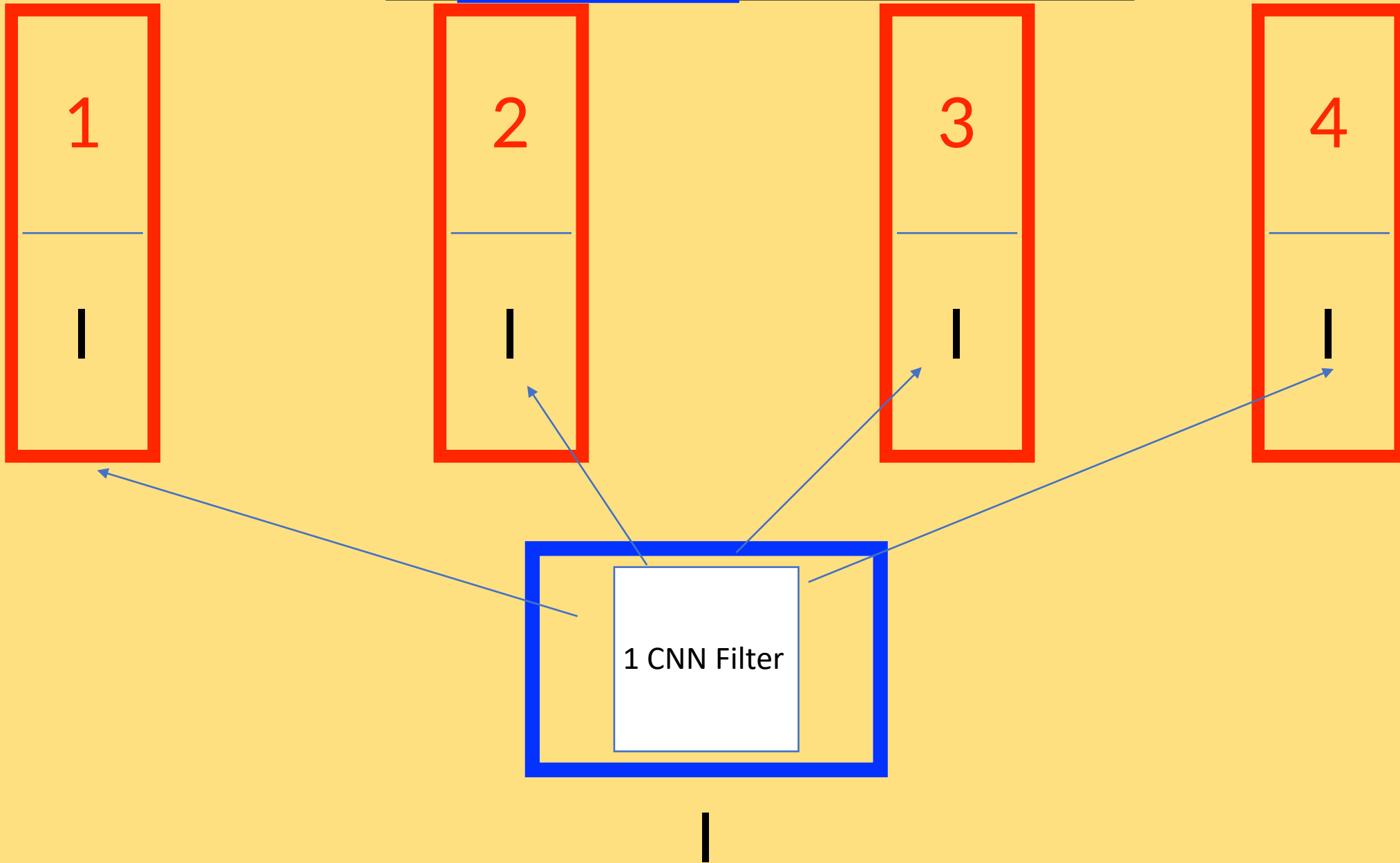


So Before FORWARD-PASS



|

So **DURING** FORWARD-PASS



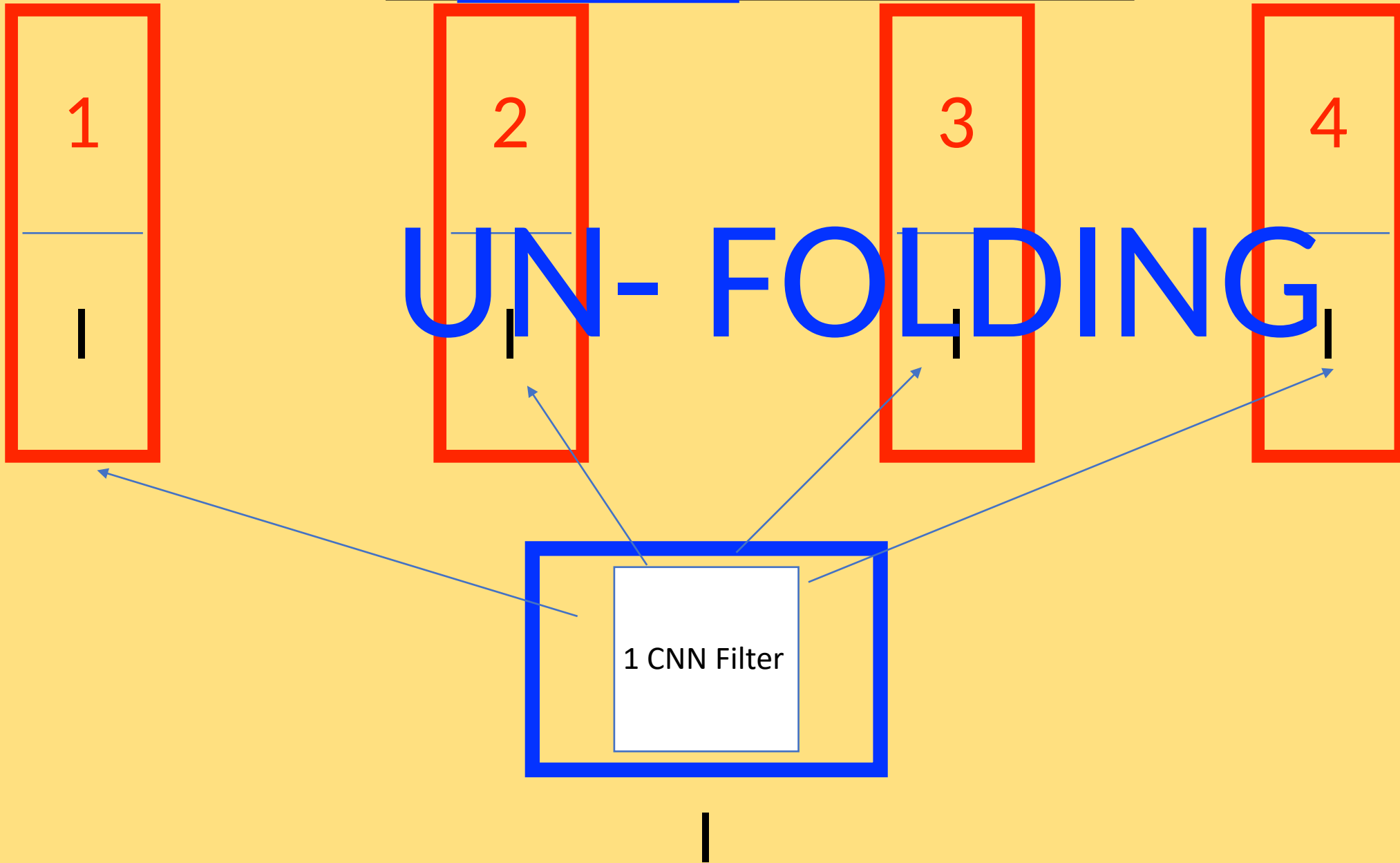
So Before FORWARD-PASS

FOLDING

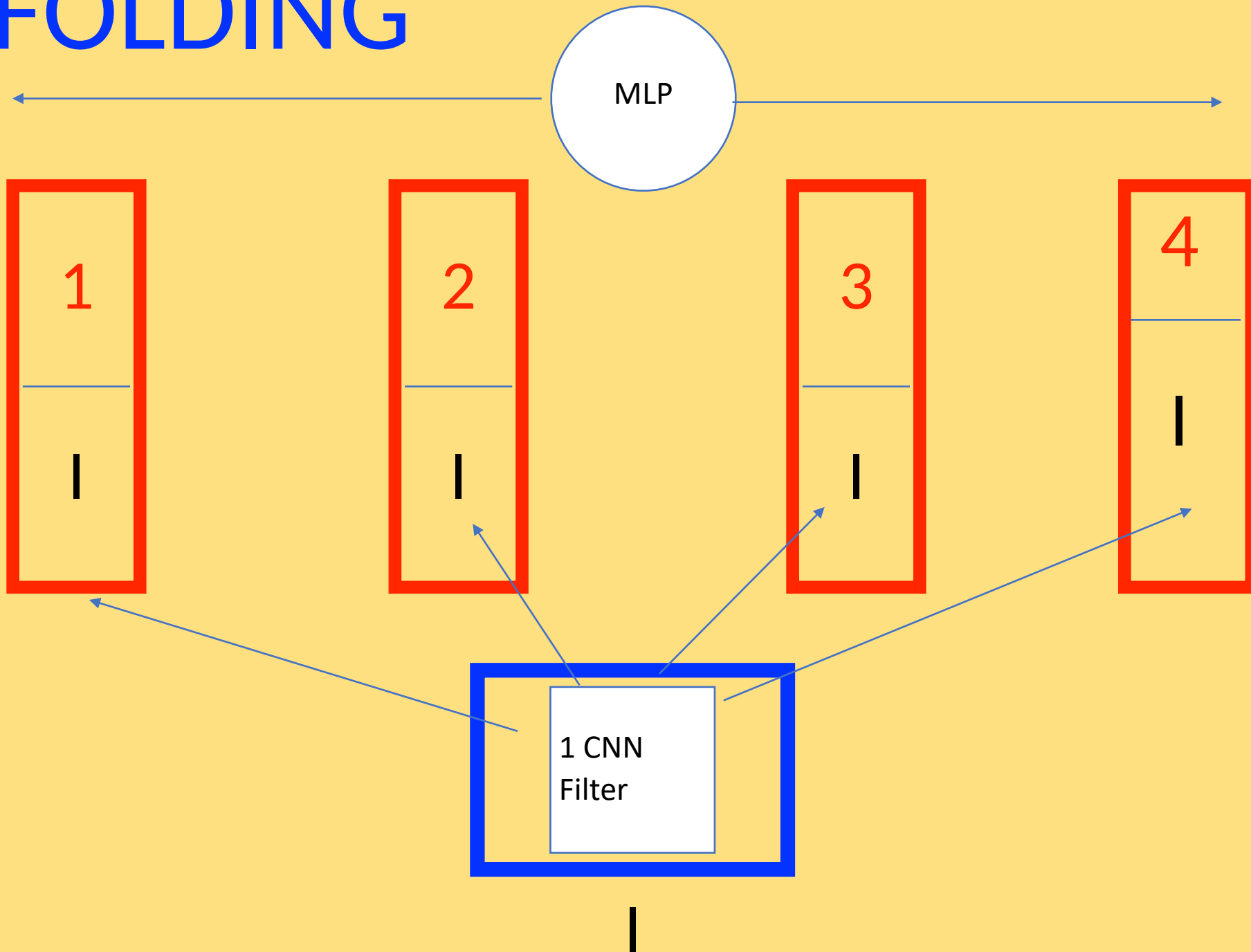


|

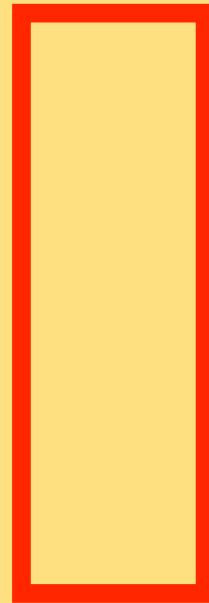
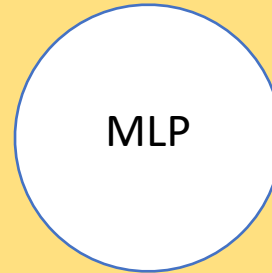
So DURING FORWARD-PASS



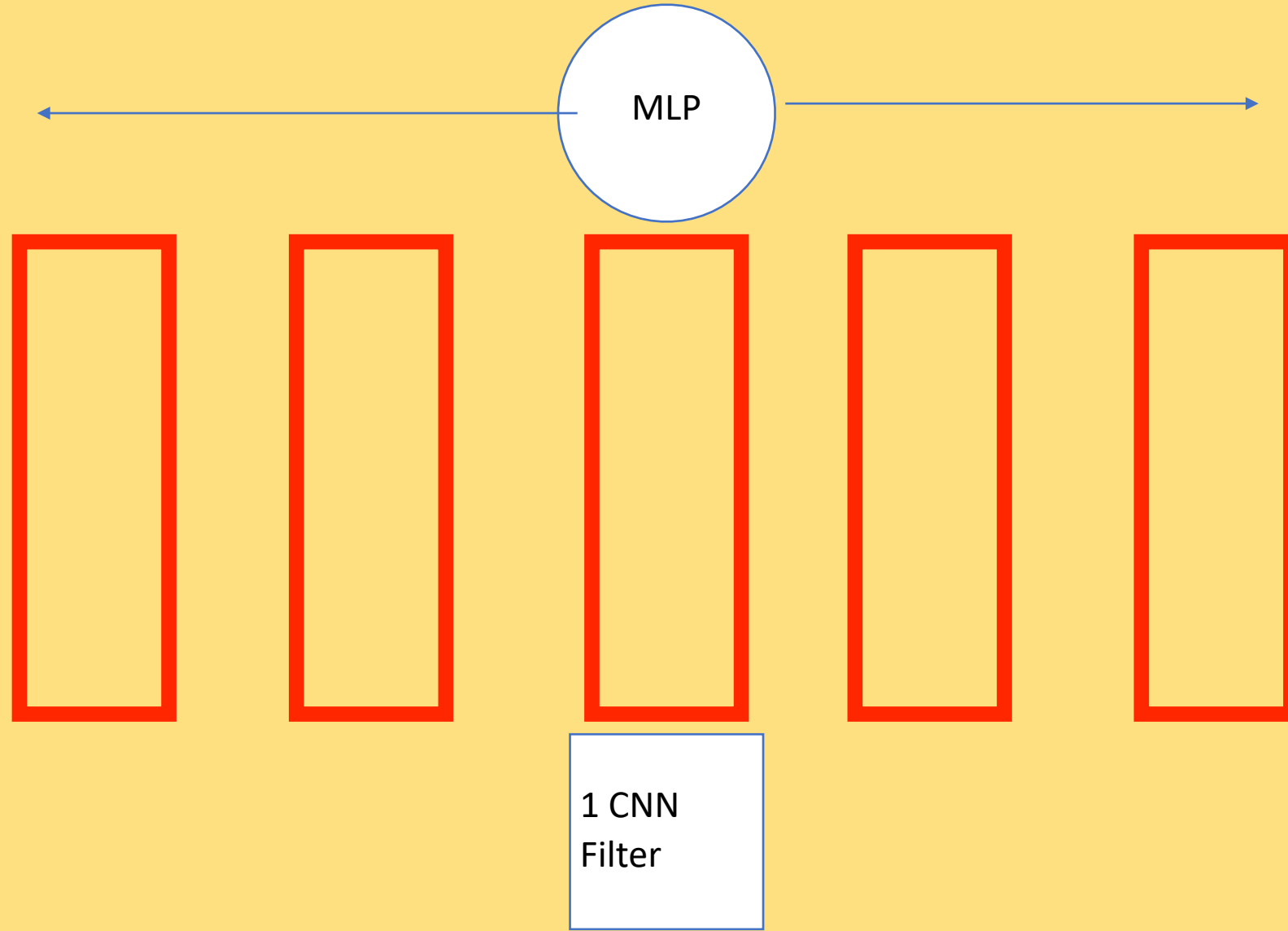
UN-FOLDING



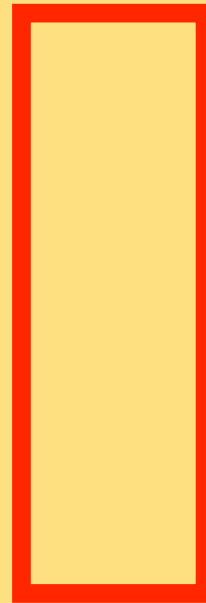
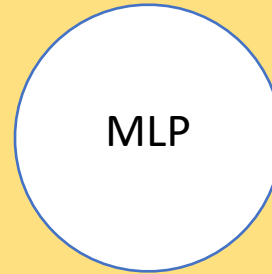
FOLDING



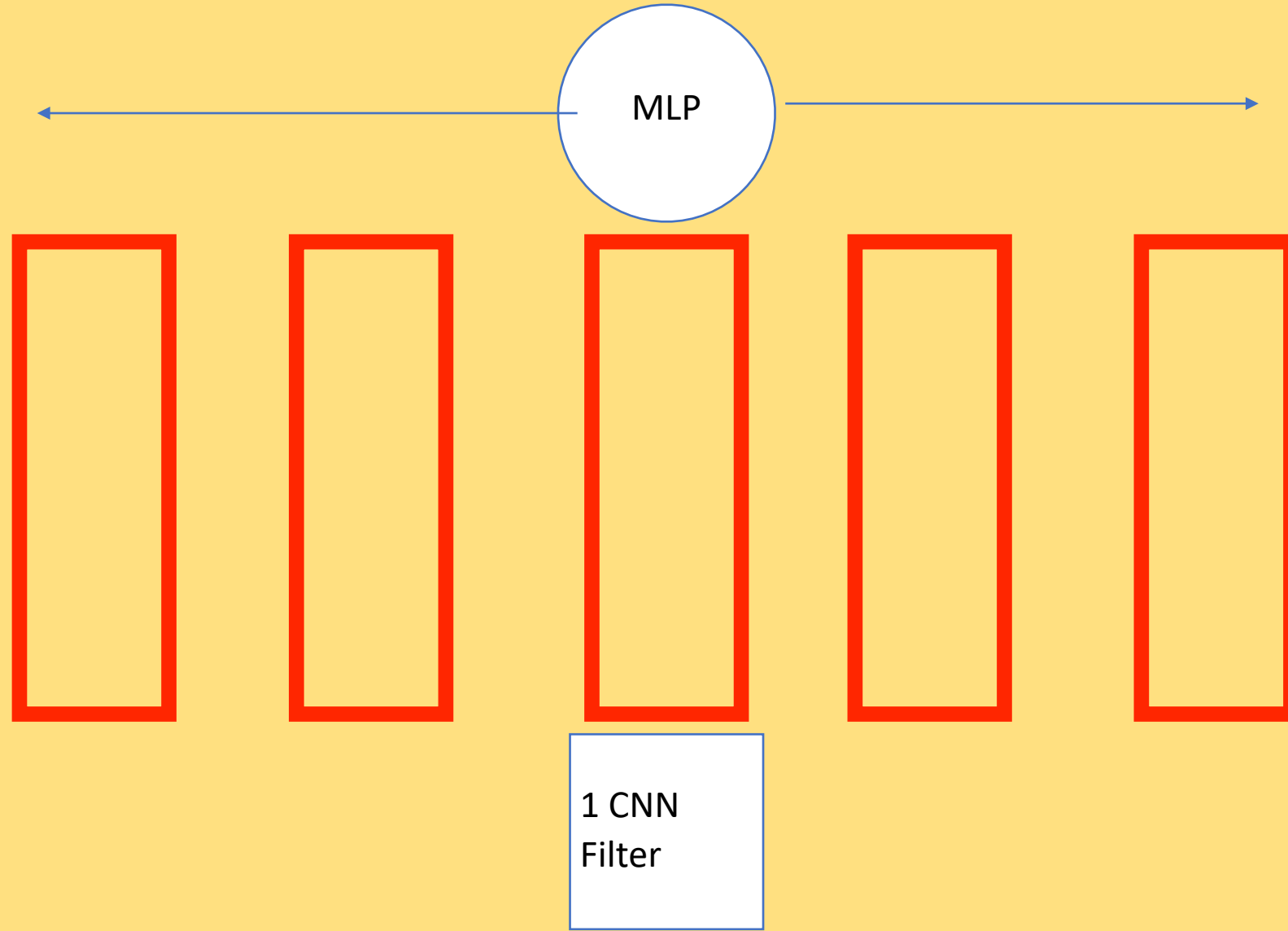
UN-FOLDING



FOLDING

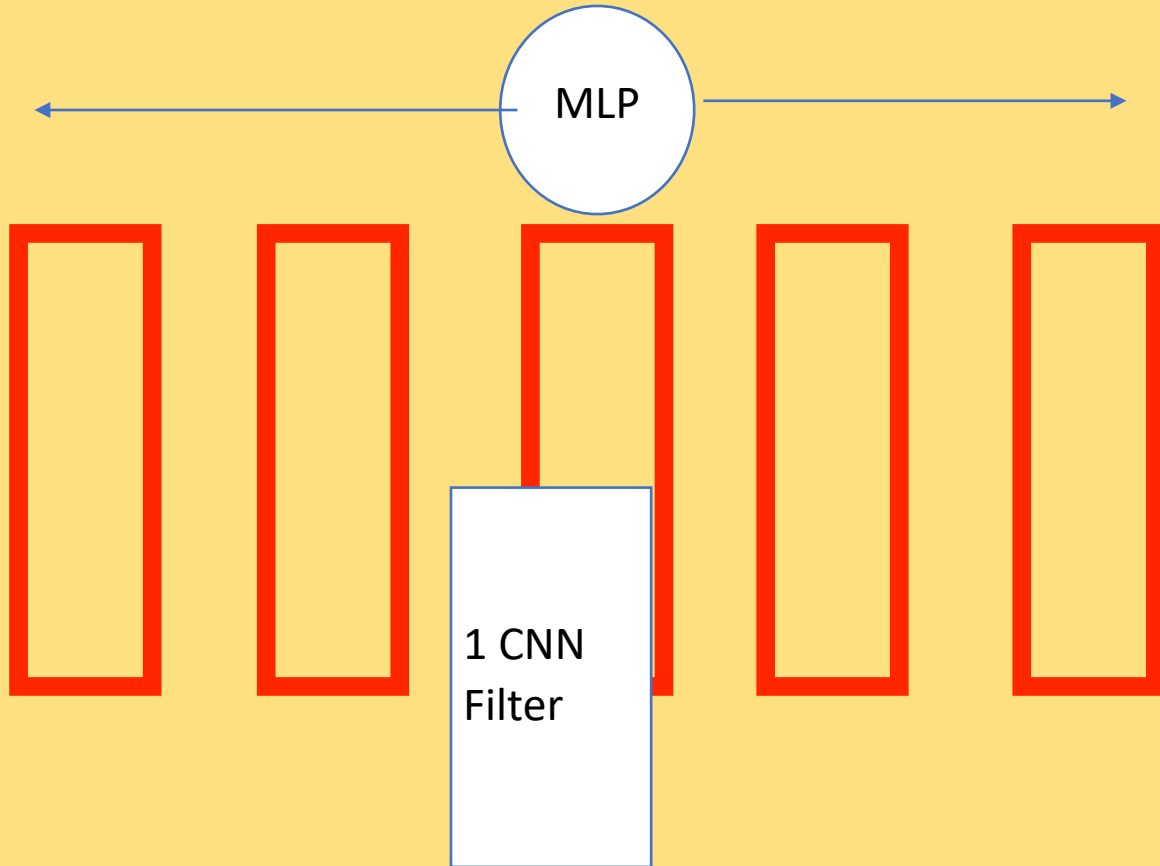


UN-FOLDING



It's that fundamental

UN-FOLDING



STONE-HENGE

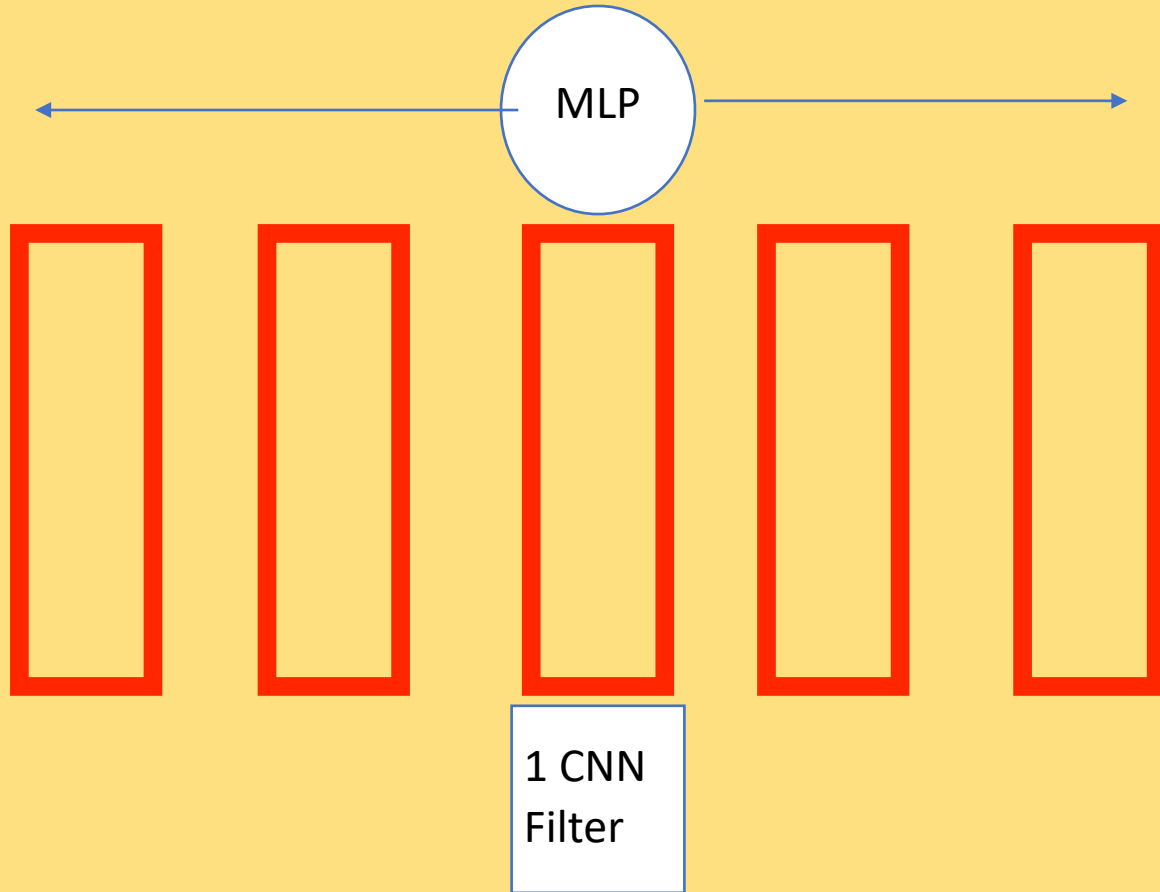


3 years.....

And this was it.

UN-FOLDING

STONEHENGE



And then I went **silent**.



[rajat] a breakthrough (hinton, nerf) and a very happy new year ➤



rajat modi <rajatmodi62@gmail.com>

to Yogesh, Yogesh ▼

📧 Sun, Dec 31, 2023, 2:53 PM

GLOM ARCHITECTURE



It's **that** fundamental

And we have to explain this
to

“peer-review.”

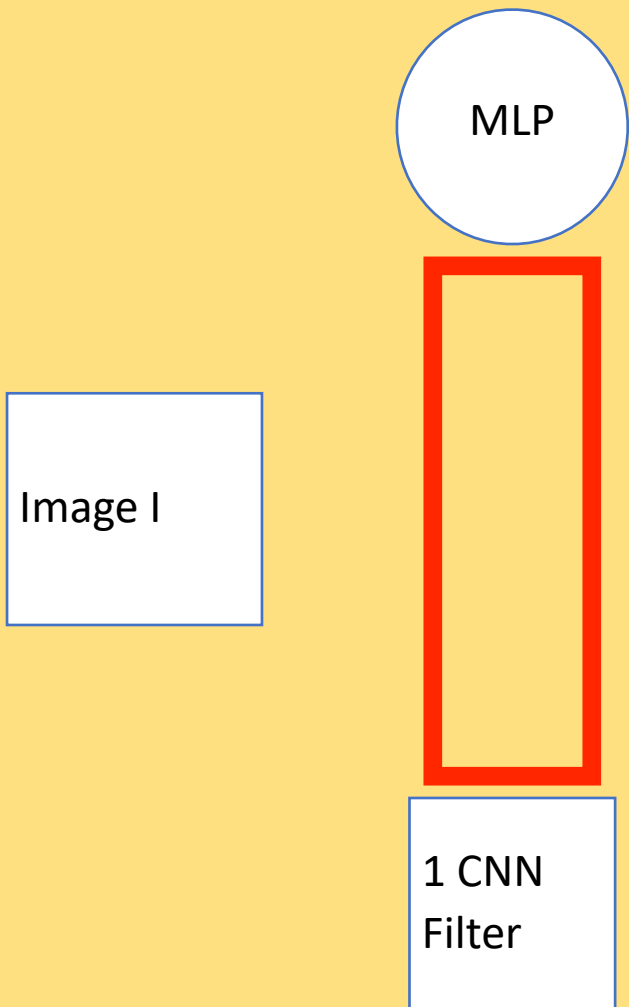
-> 10000 reviewers.

-> and hope someone's mind
is open enough to see this.

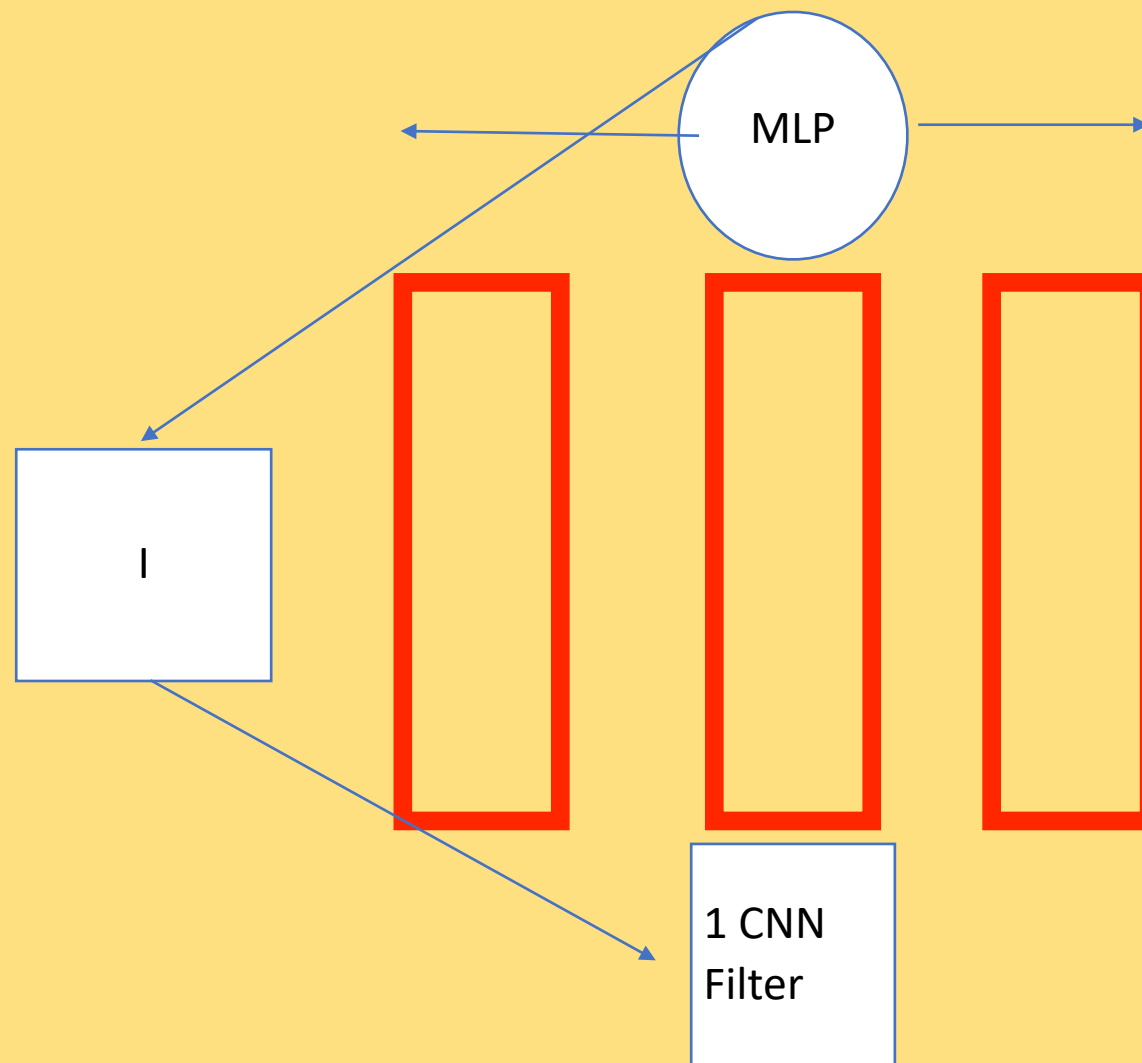
-> 2/4 reviewers dont even
understand it

How to **train** GLOM

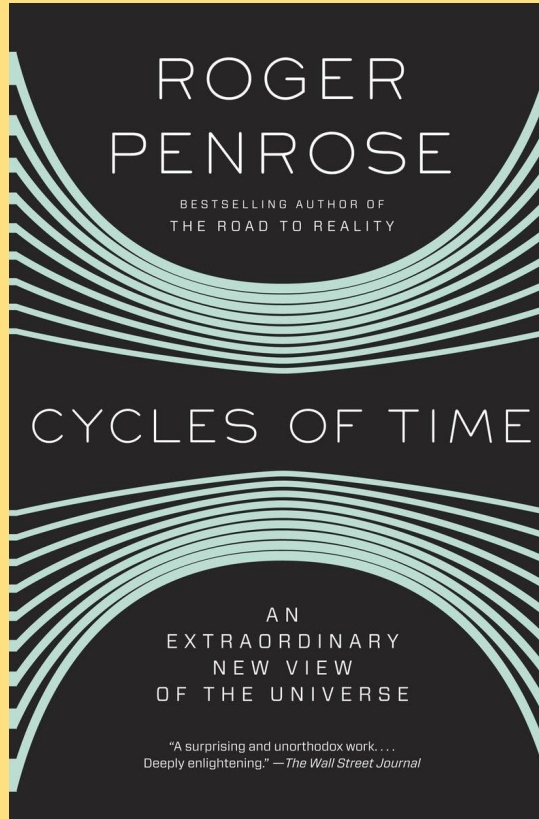
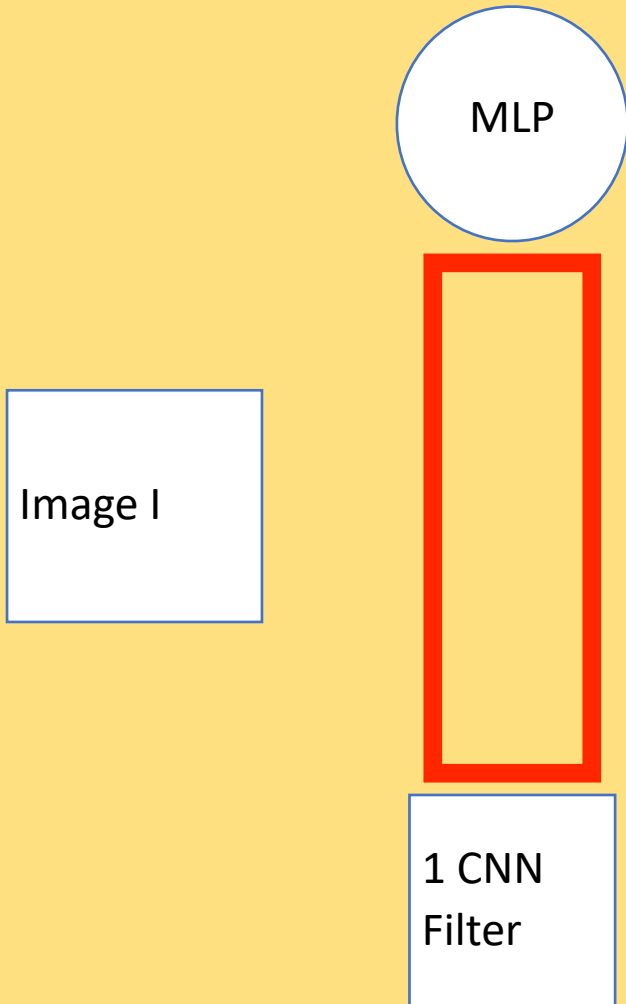
FOLDED



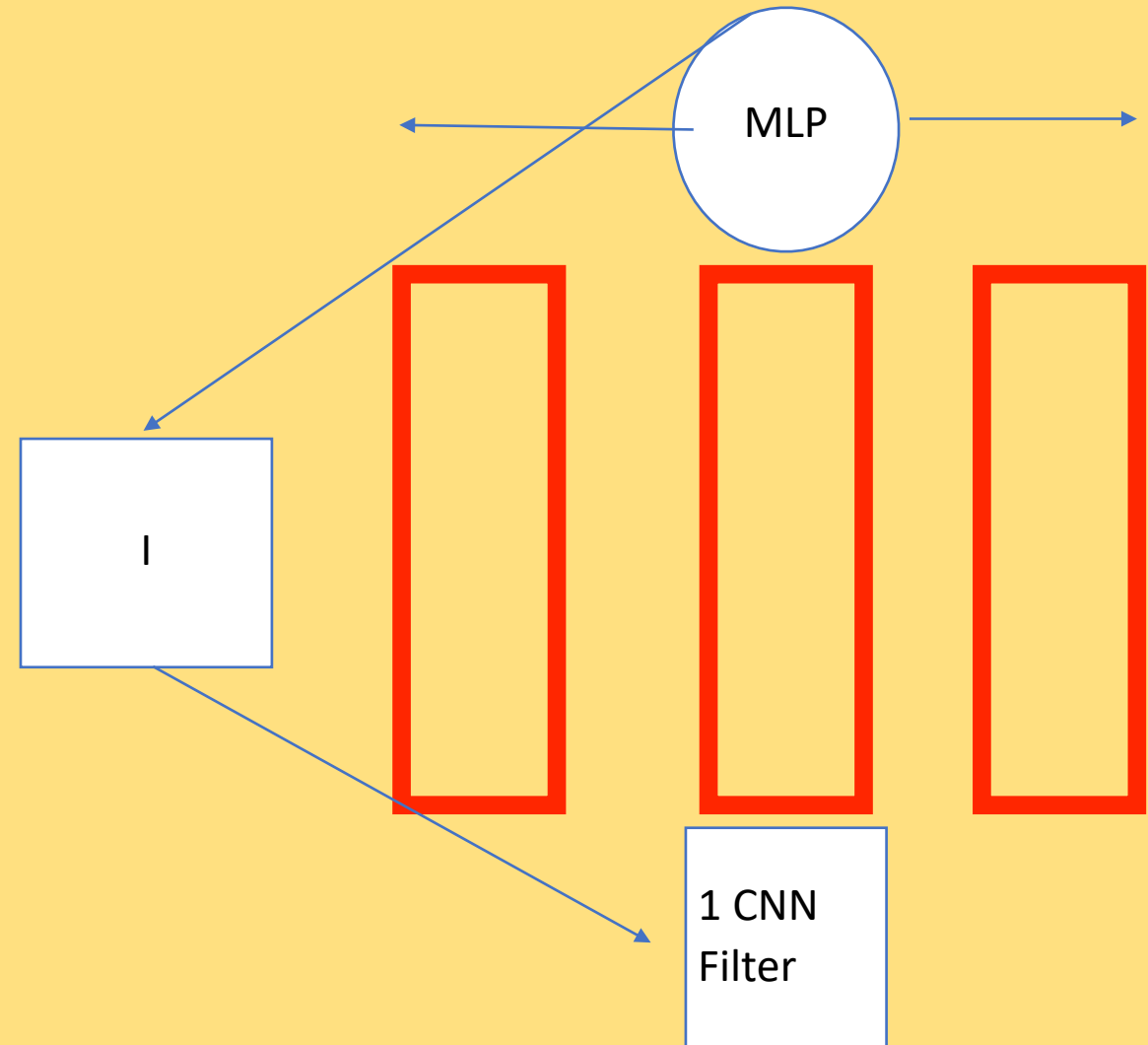
UN-FOLDED



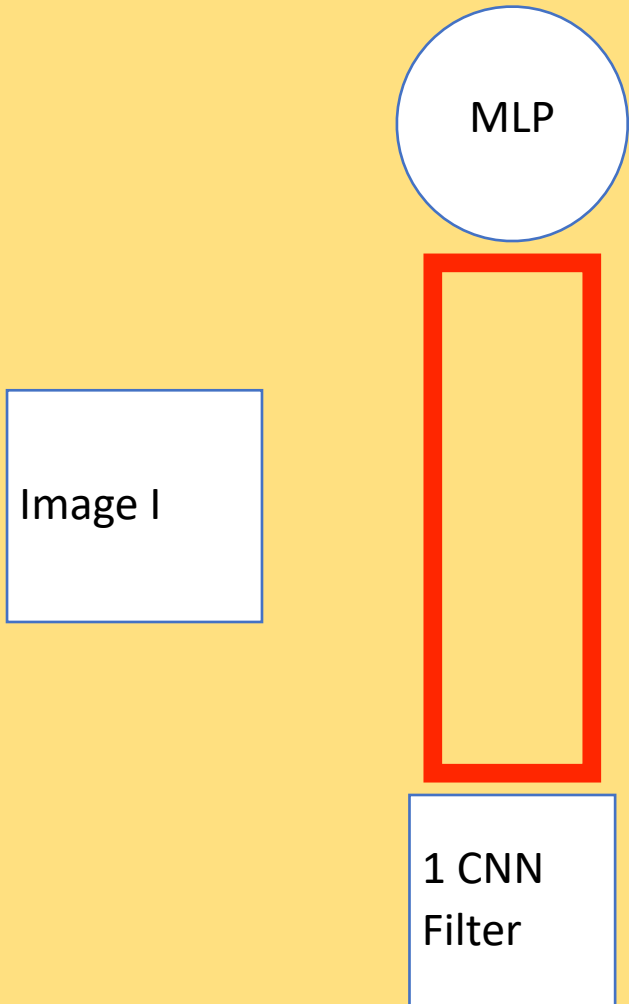
FOLDED



UN-FOLDED

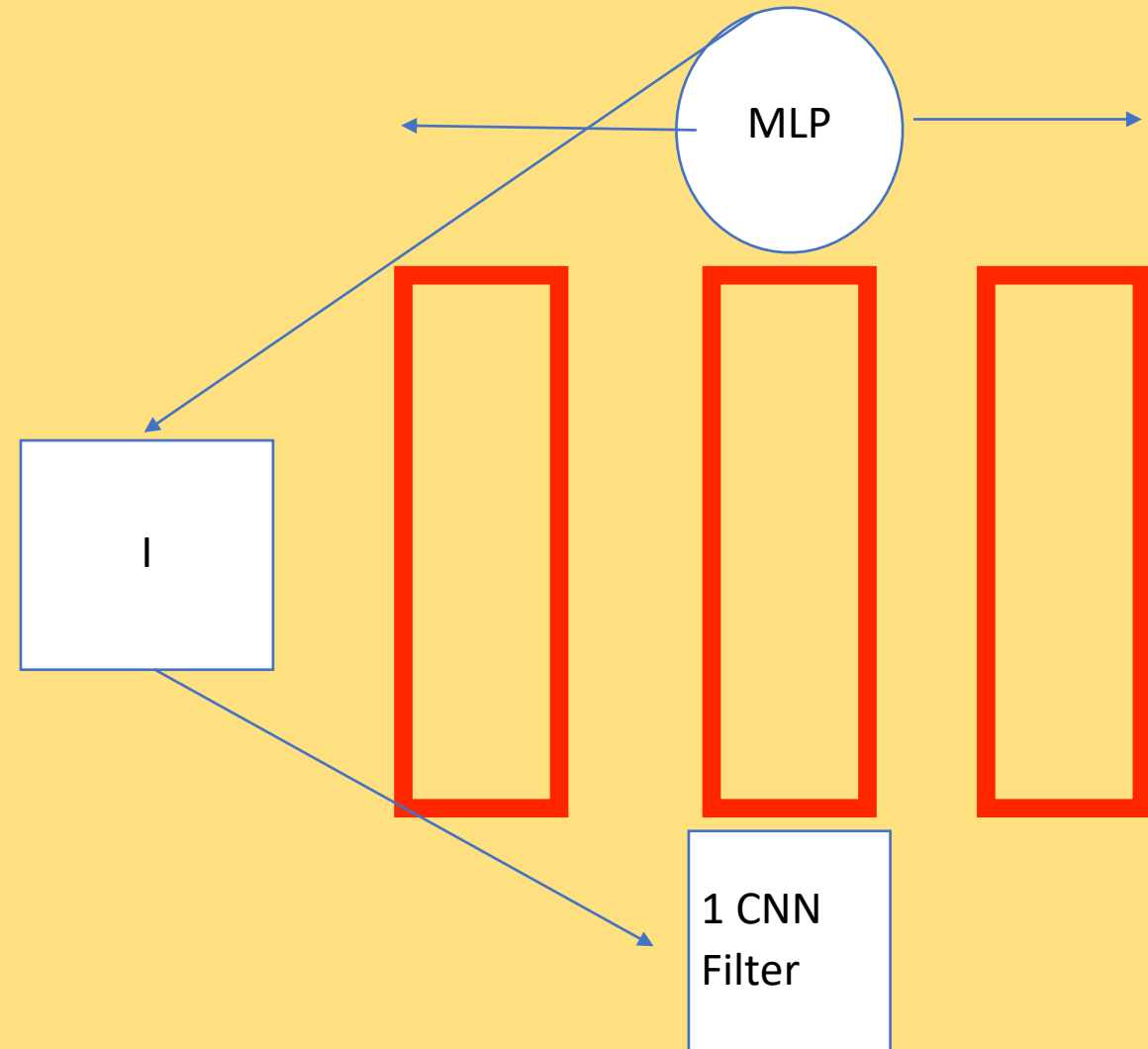


FOLDED

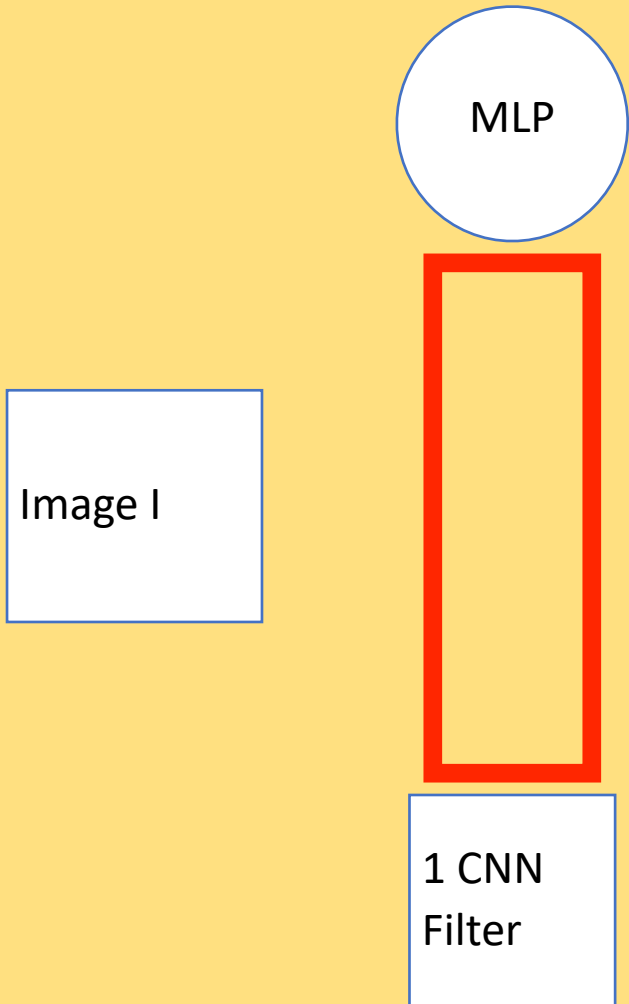


Cycles of learning iterations

UN-FOLDED

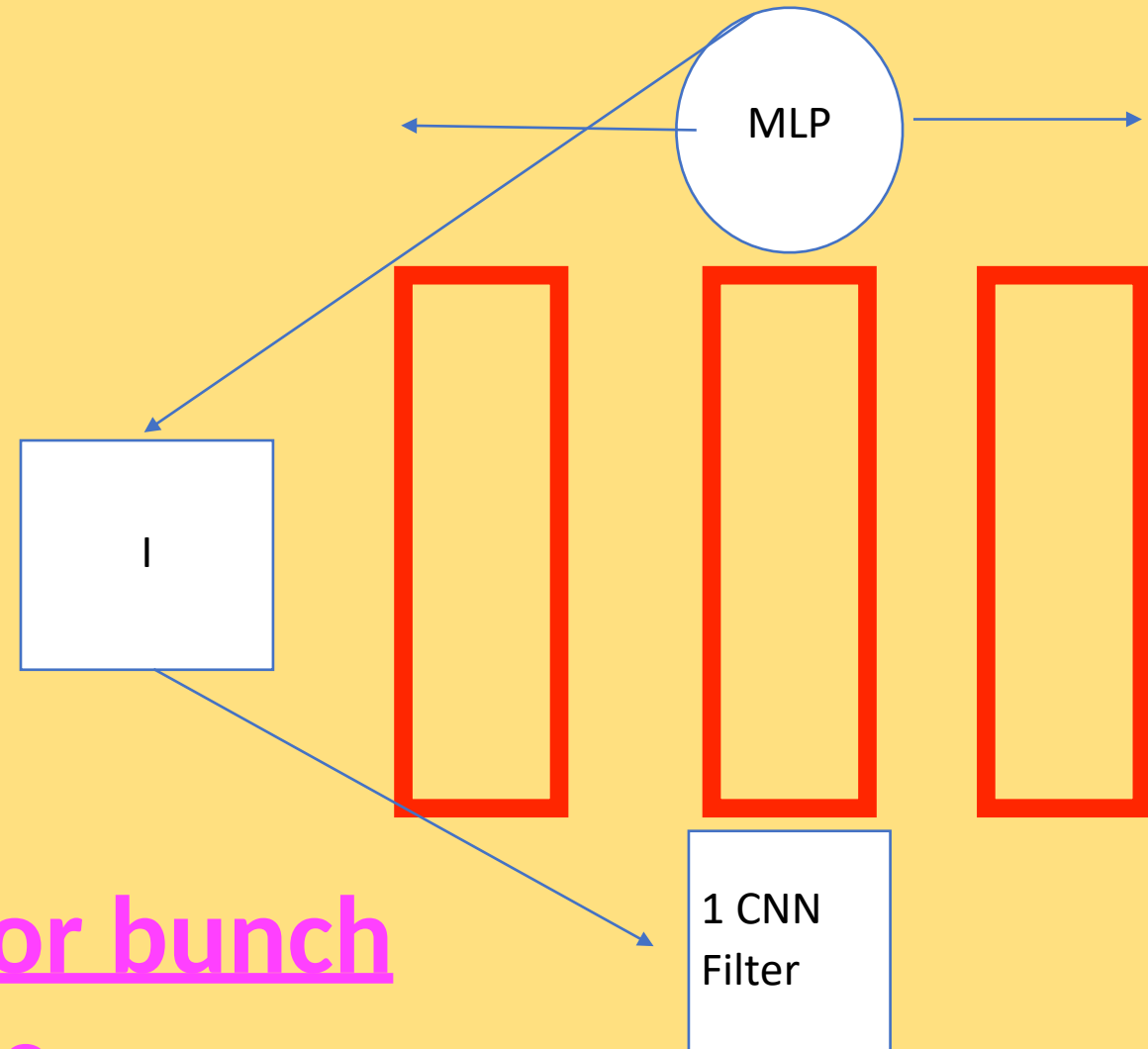


FOLDED



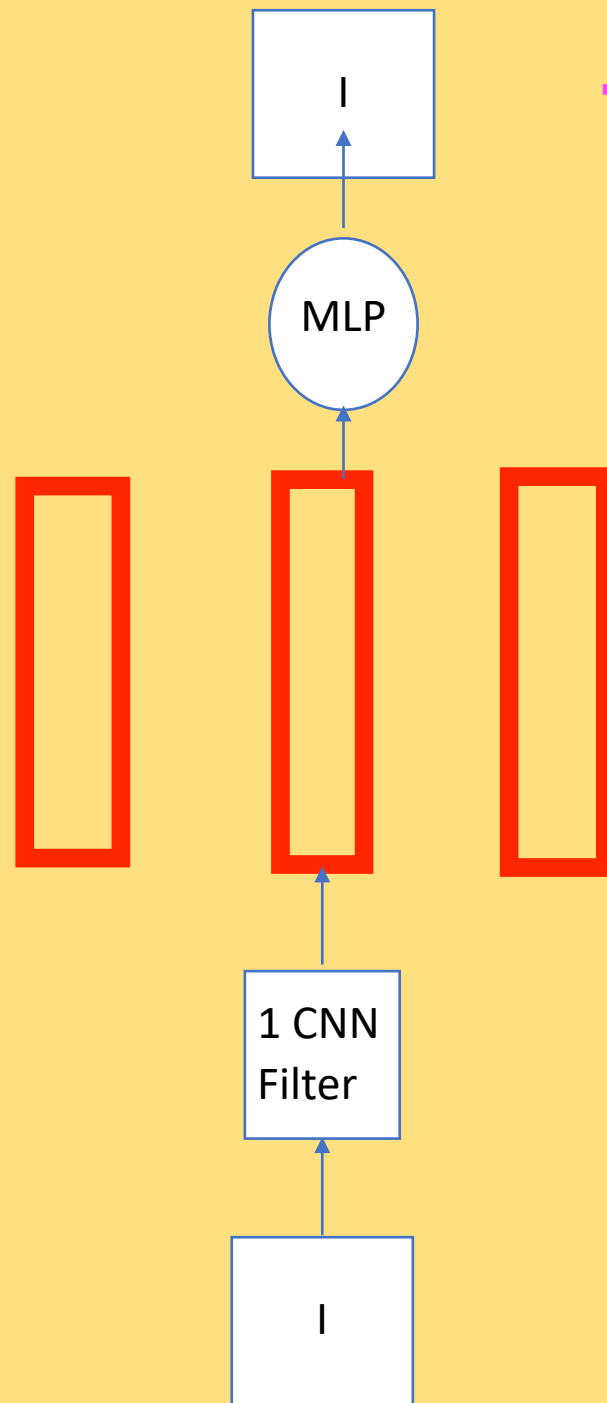
Cycles of learning iterations

UN-FOLDED



Do this for bunch of images.

During Inference



UN-FOLDED

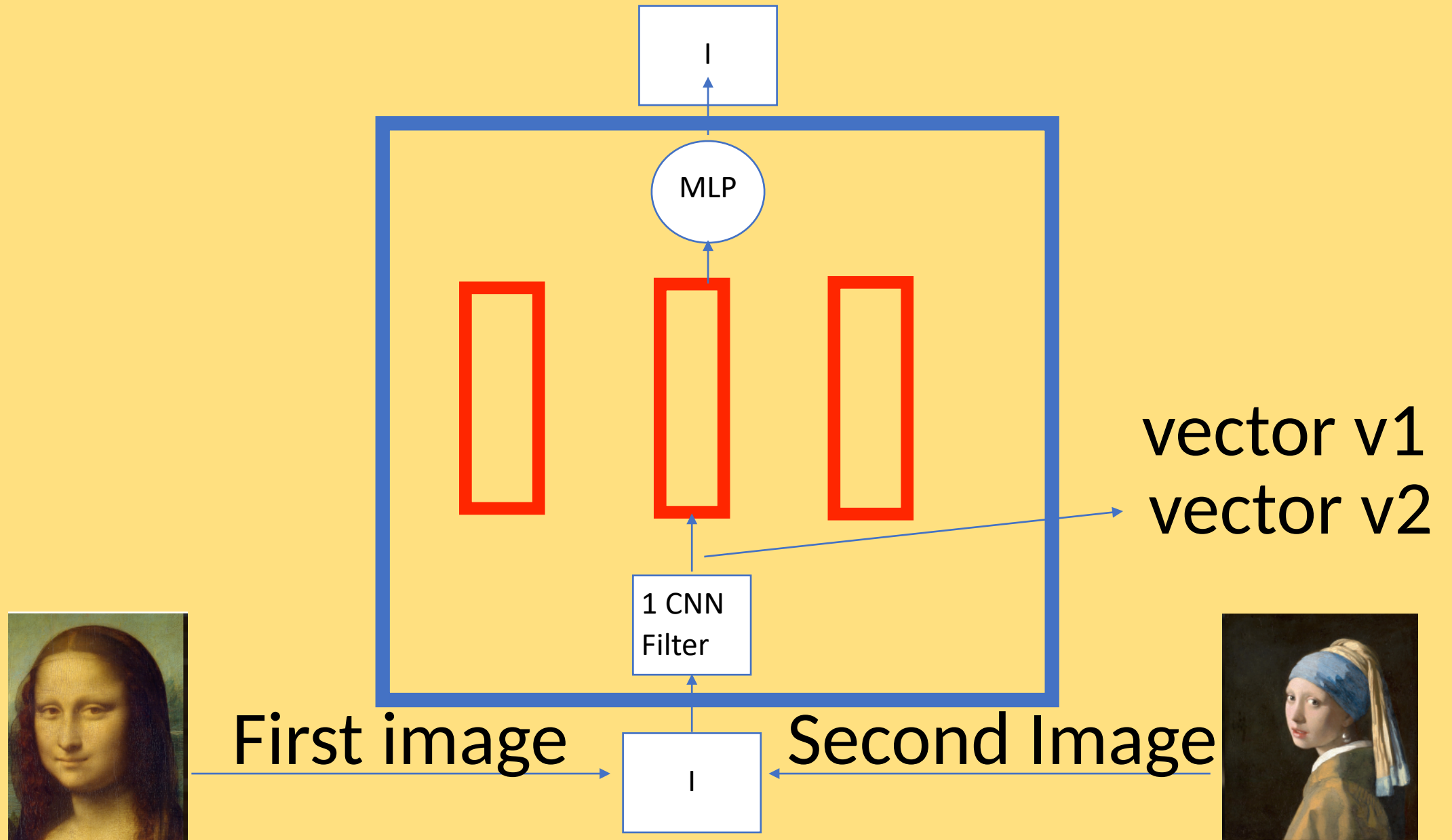
SO what ? You **Feed** the same image in and **get it back.**

-> You **just did** it with **MLP.**

-> **MAE** did it with a **transformer.**

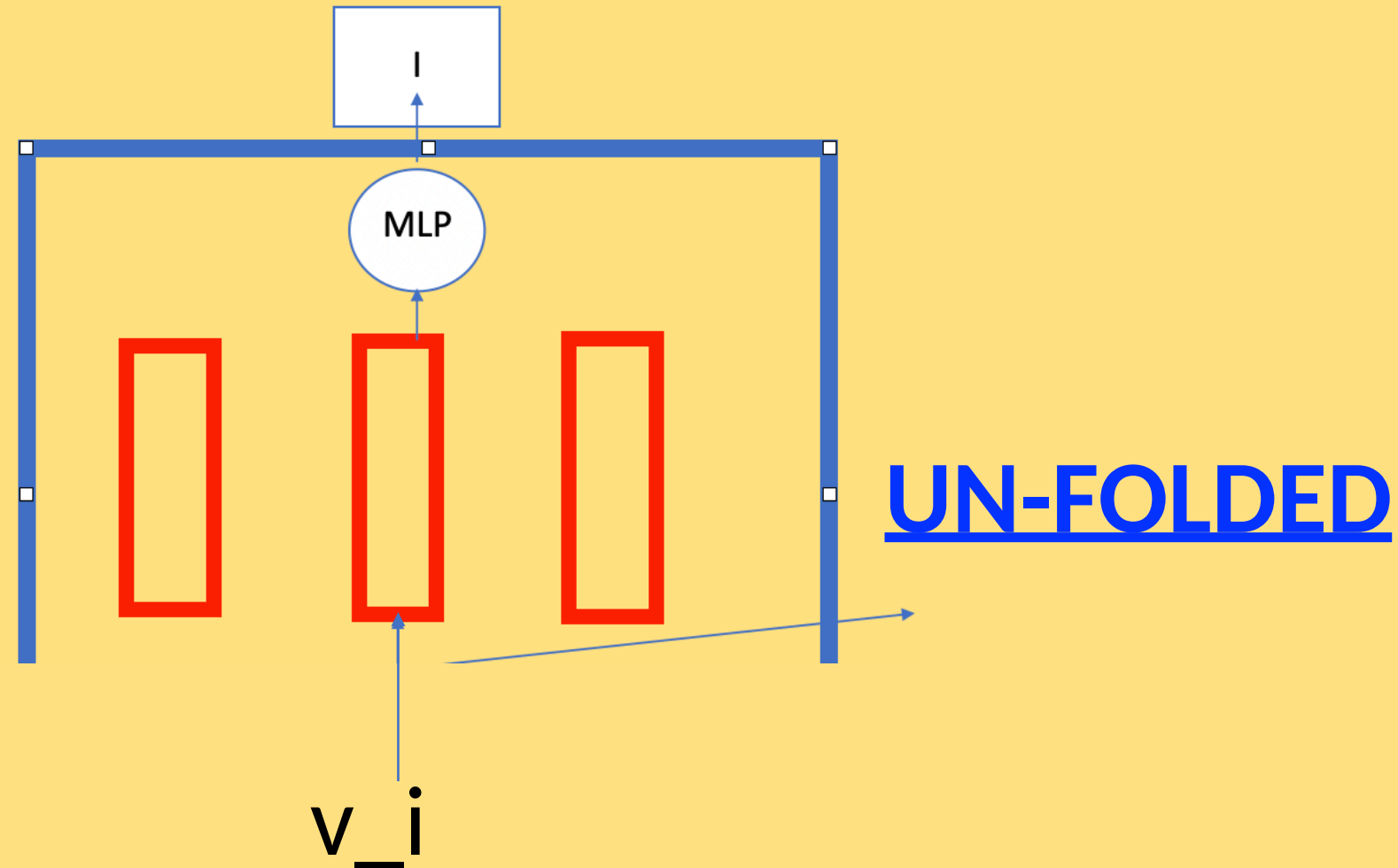
-> How is it **different** from Masked Auto-encoder?

-> How is it any **different** from **MAE**?

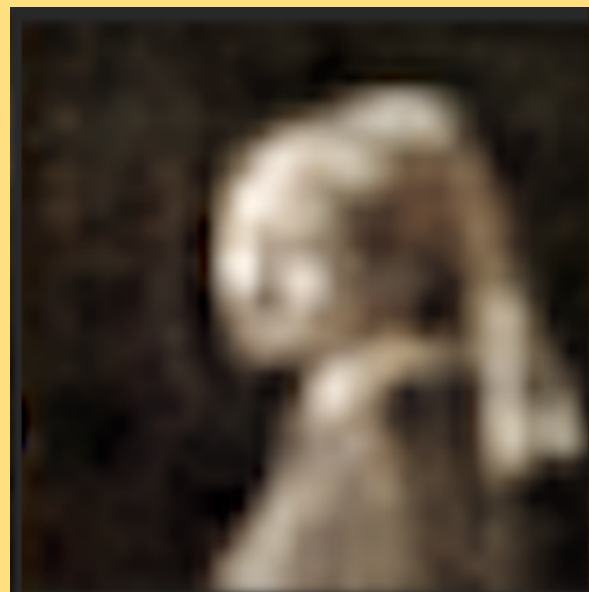


vector v1 vector v2

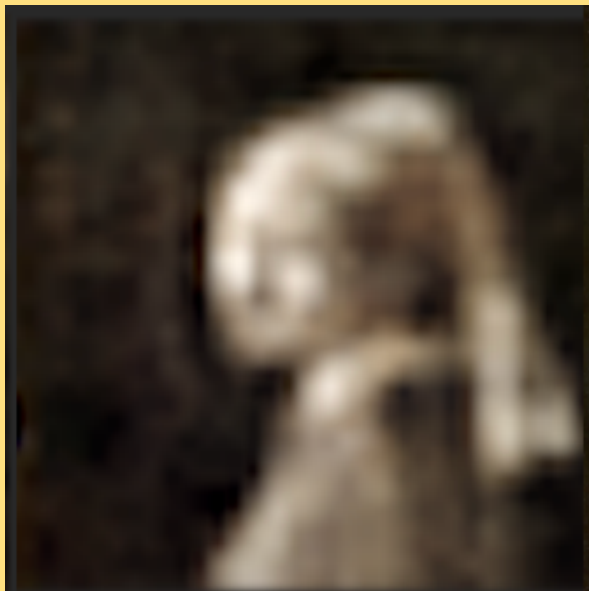
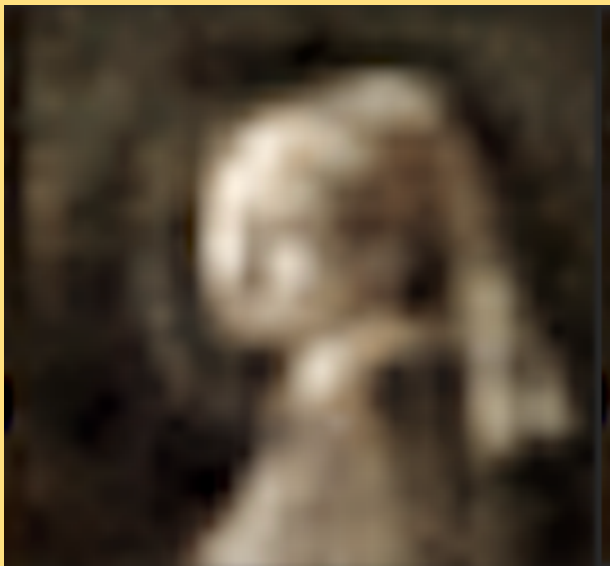
$$\text{vector } v_i = v1 + (v2-v1)/n$$



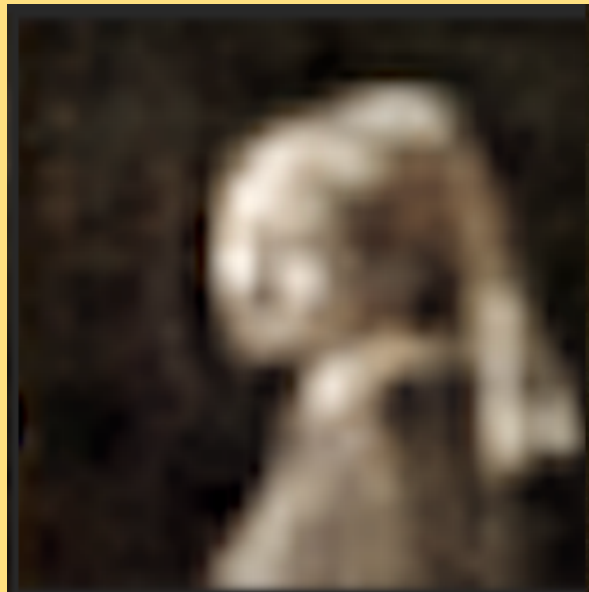
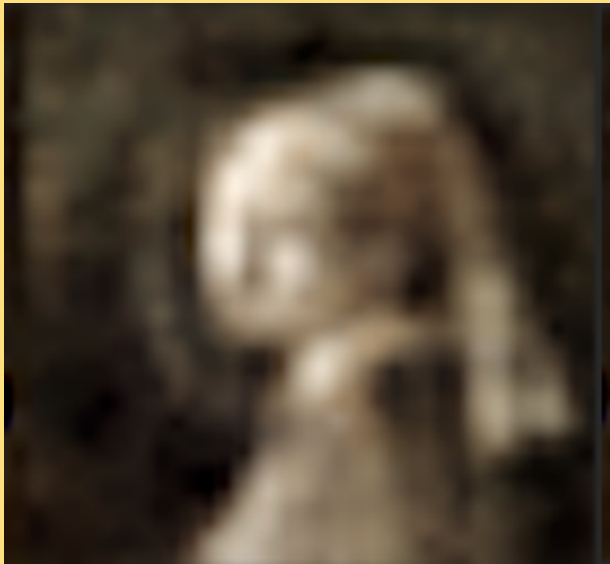
THIS IS WHAT YOU GET

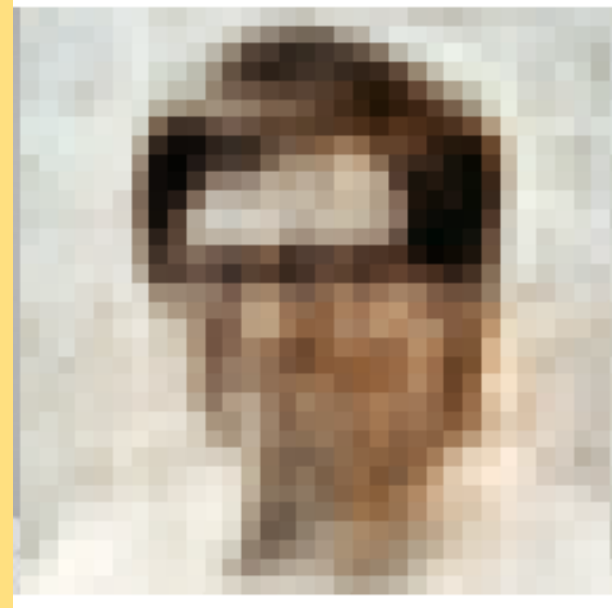


YOU CAN INTERPOLATE. NO MORE COLLAPSE.



JUST FOLDING-UNFOLDING.





Black and White

Becomes Colored

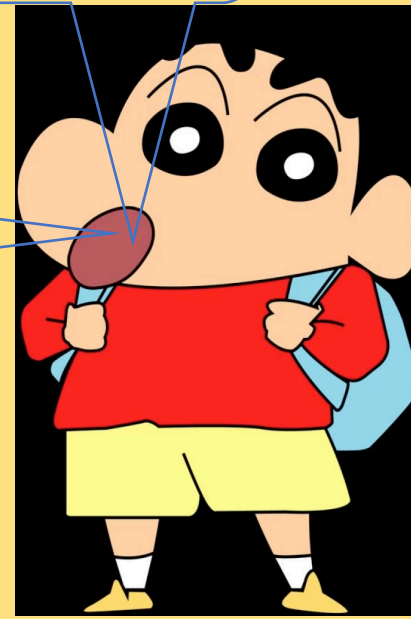
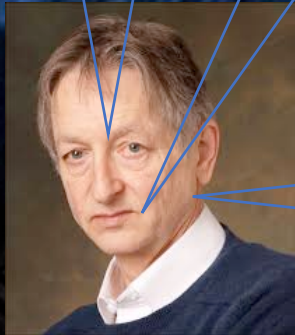
Geez kiddo,
Always **impatient**,
Not yet,

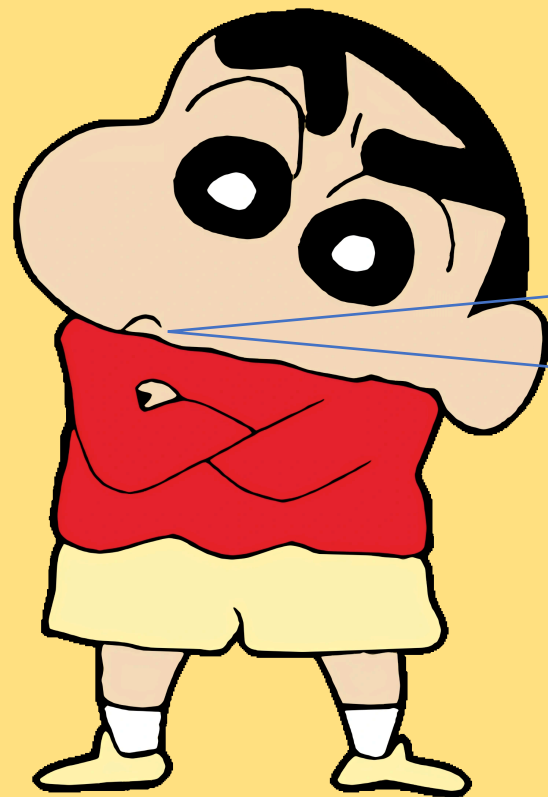
Train Time?

Geoff sir,
Geoff sir,
MLP does image-reconstruction
now,
My school-teacher told me it can't
Can we make paper?
ChatGPT can write it.

Ewwwww!!!!,
Geoff

1-2 days geoff sir, as
usual....





Ewww says Geoff
Need to take care of this,
GLOMMMM,

Steal Another idea.....

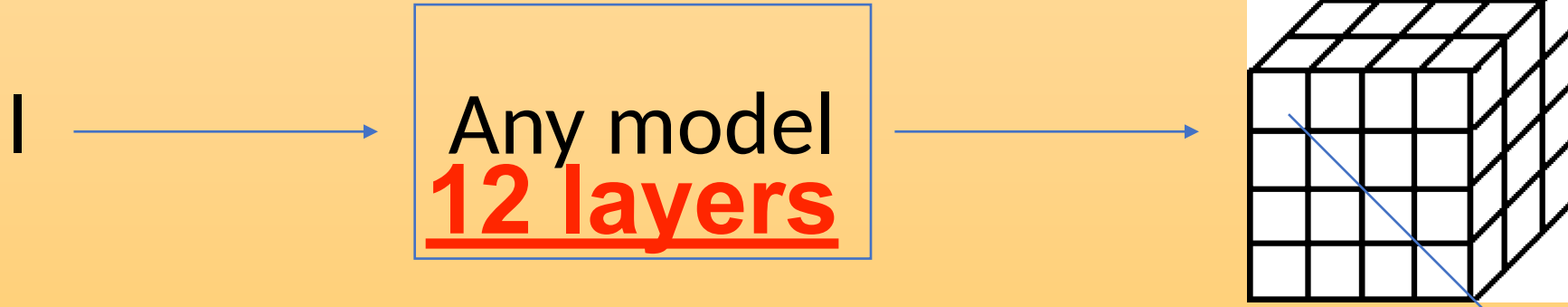


One advantage of sharing knowledge between locations via distillation rather than by copying weights is that the inputs to the bottom-up models at different locations do not need to have the same structure. This makes it easy to have a retina whose receptive fields get progressively larger further from the fovea, which is hard to handle using weight-sharing in a convolutional net. Many other aspects, such as the increase in chromatic aberration further from the fovea are also easily handled. Two corresponding nets at different locations should learn to compute the same function of the optic array even though this array is pre-processed differently by the imaging process before being presented to the two nets. Co-distillation also means that the top-down models do not need to receive their location as an input since it is always the same for any given model.

- How to make it even faster?

Layer Skipping

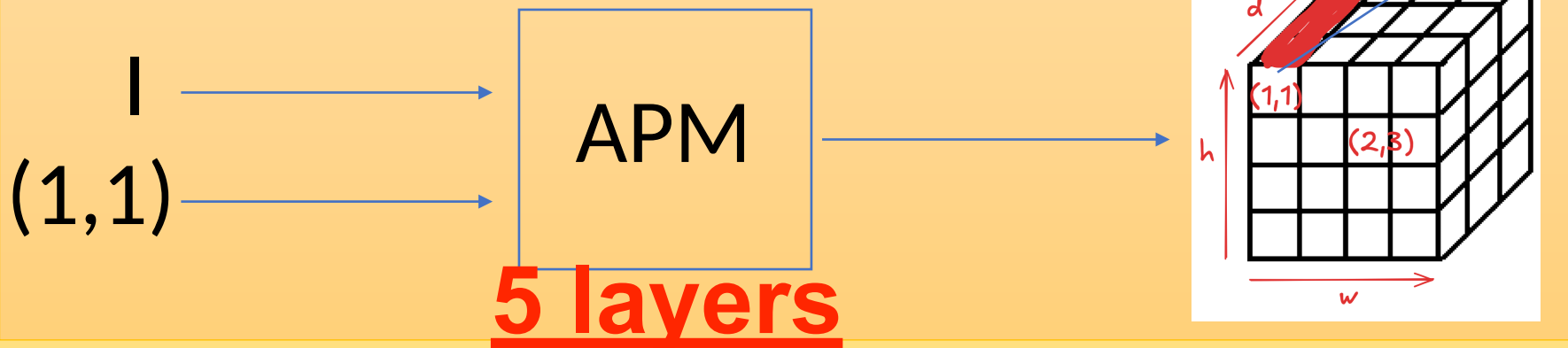
Parallel Perception



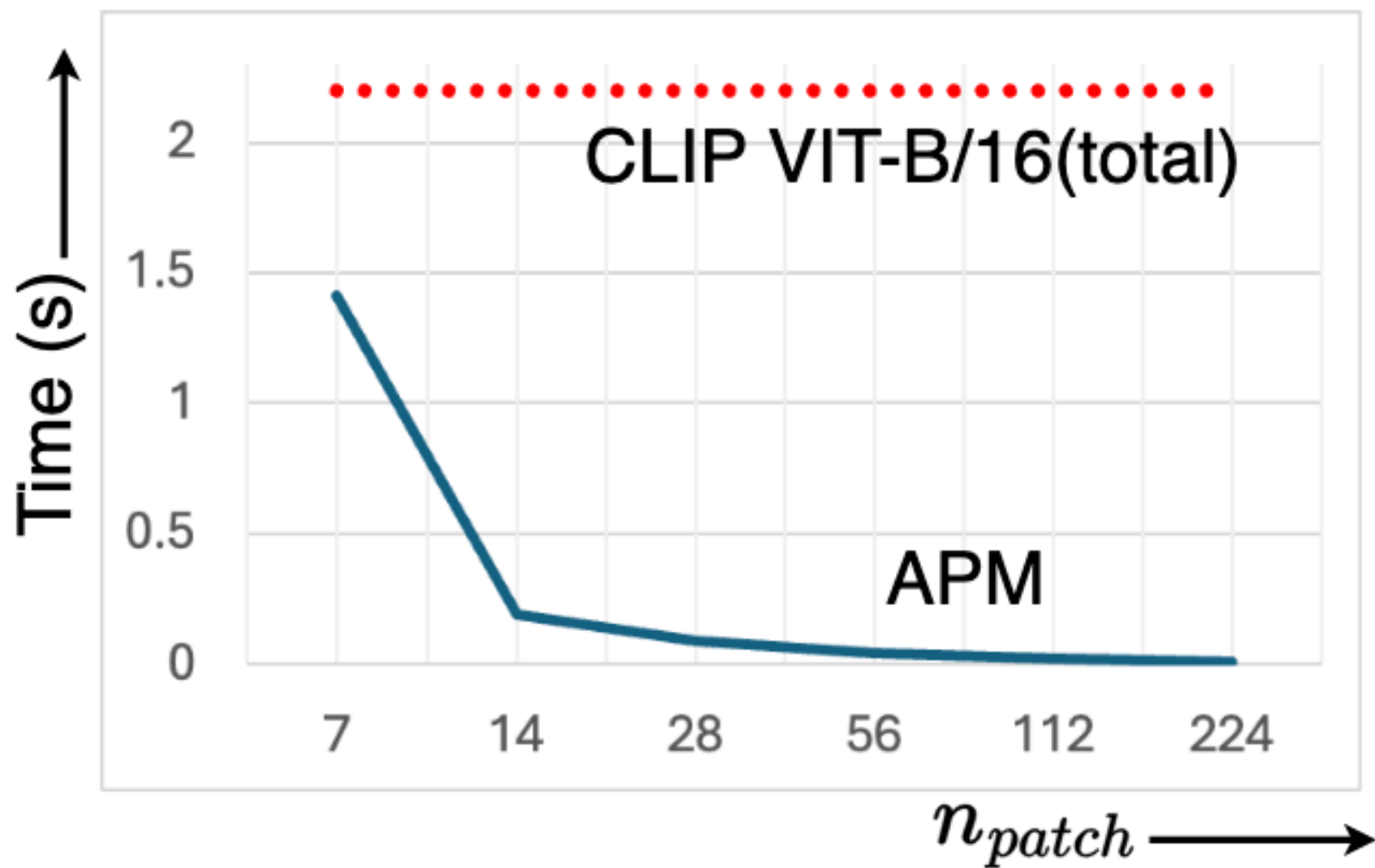
Distillation

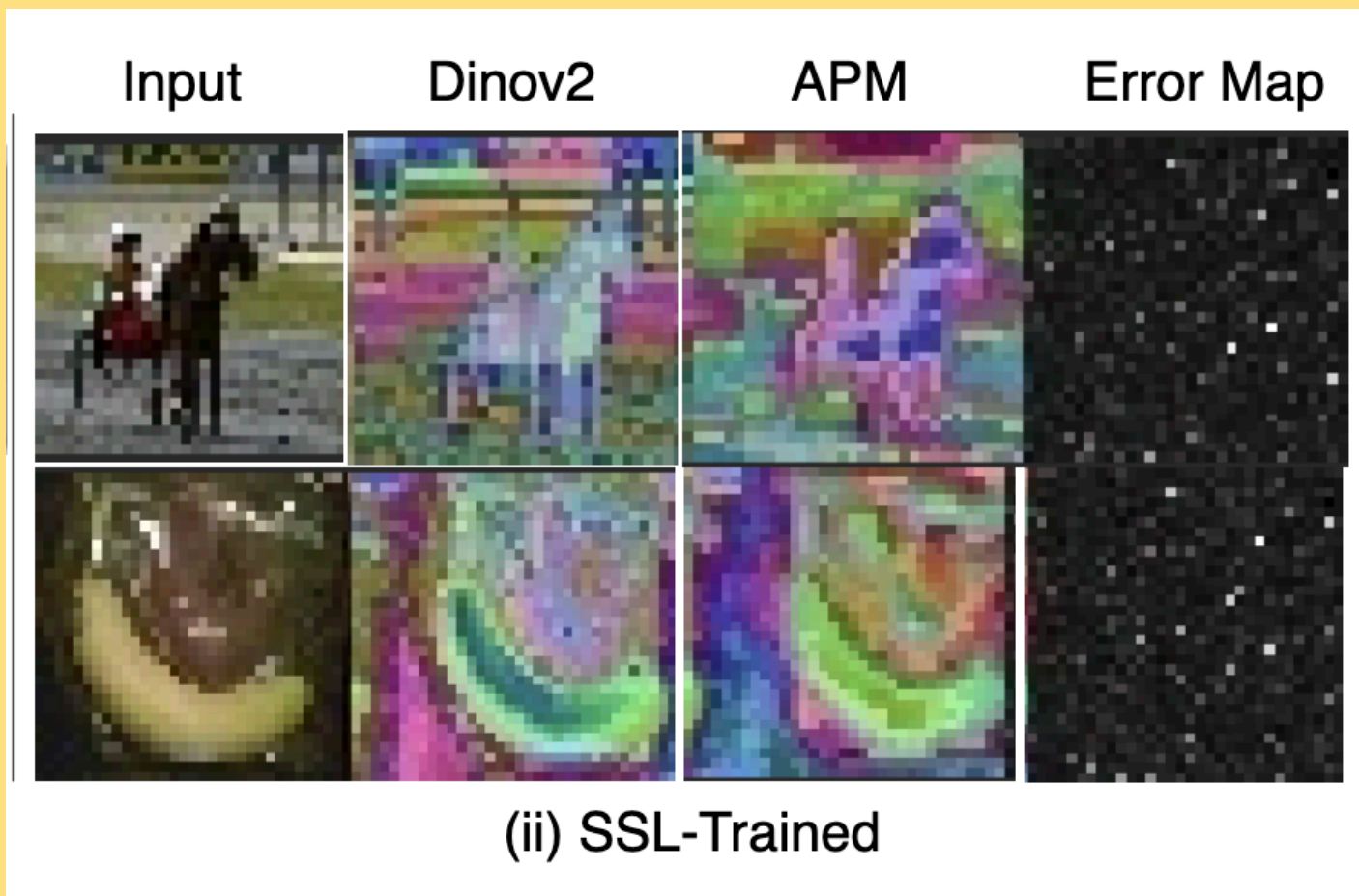


Asynchronous Perception



Inference Time vs No of Patches



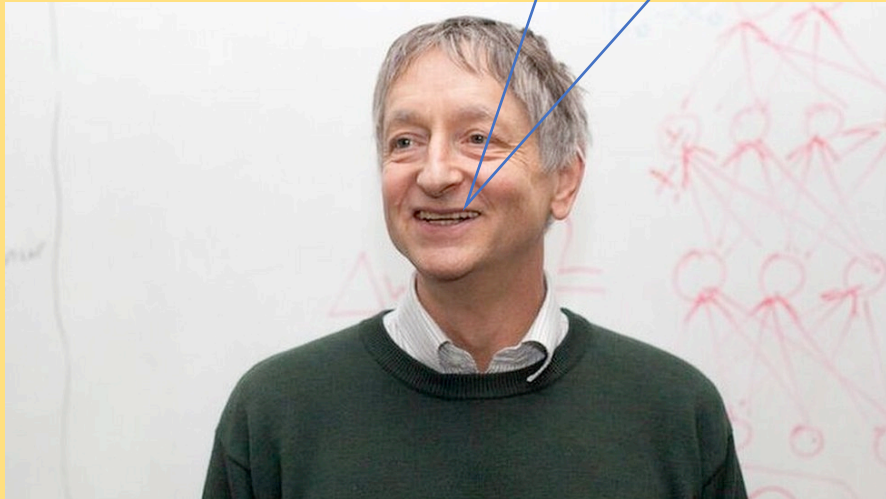


Training?

Eww,

Ewwww,

EWWWWWW,



Geoff sir,

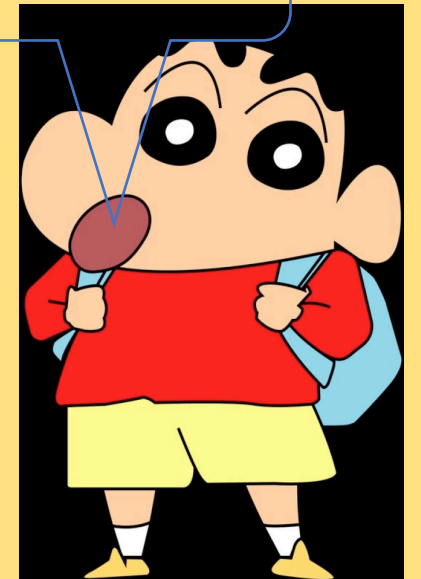
Geoff sir,

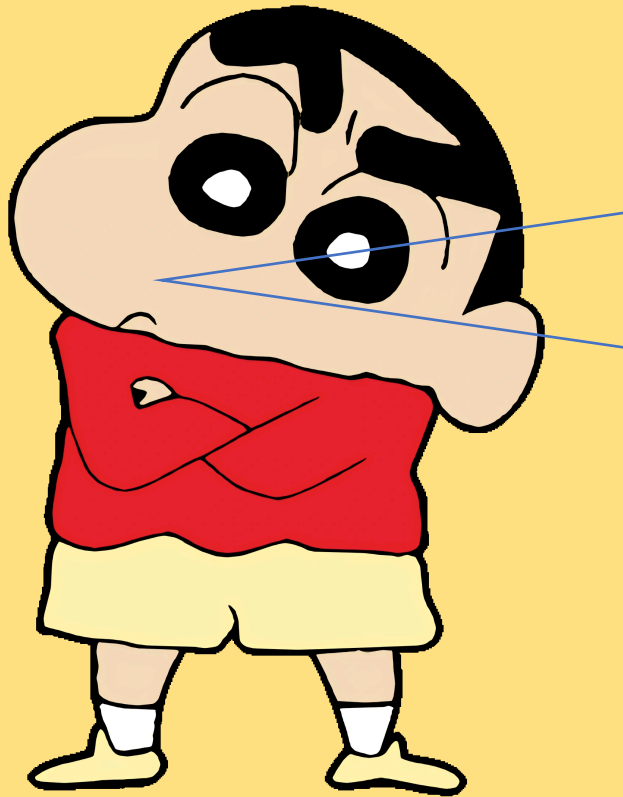
It trains in 2 hours

instead of a day,

Can we make a paper

out of this?





Ewww says Geoff.....**THRICE** this
time,
Need to **take care** of this,
GLOMMMMM,

DON't use many samples

Currently, we do not exploit this interesting property of FF because we still use mini-batches, but the ability of a deep neural net to absorb a lot of information from a **single** training case by jumping to a set of weights that handles that case perfectly could be of interest to psychologists who are tired of creeping down gradients²⁰

Just use **1 sample.**

Just use **1 sample.**

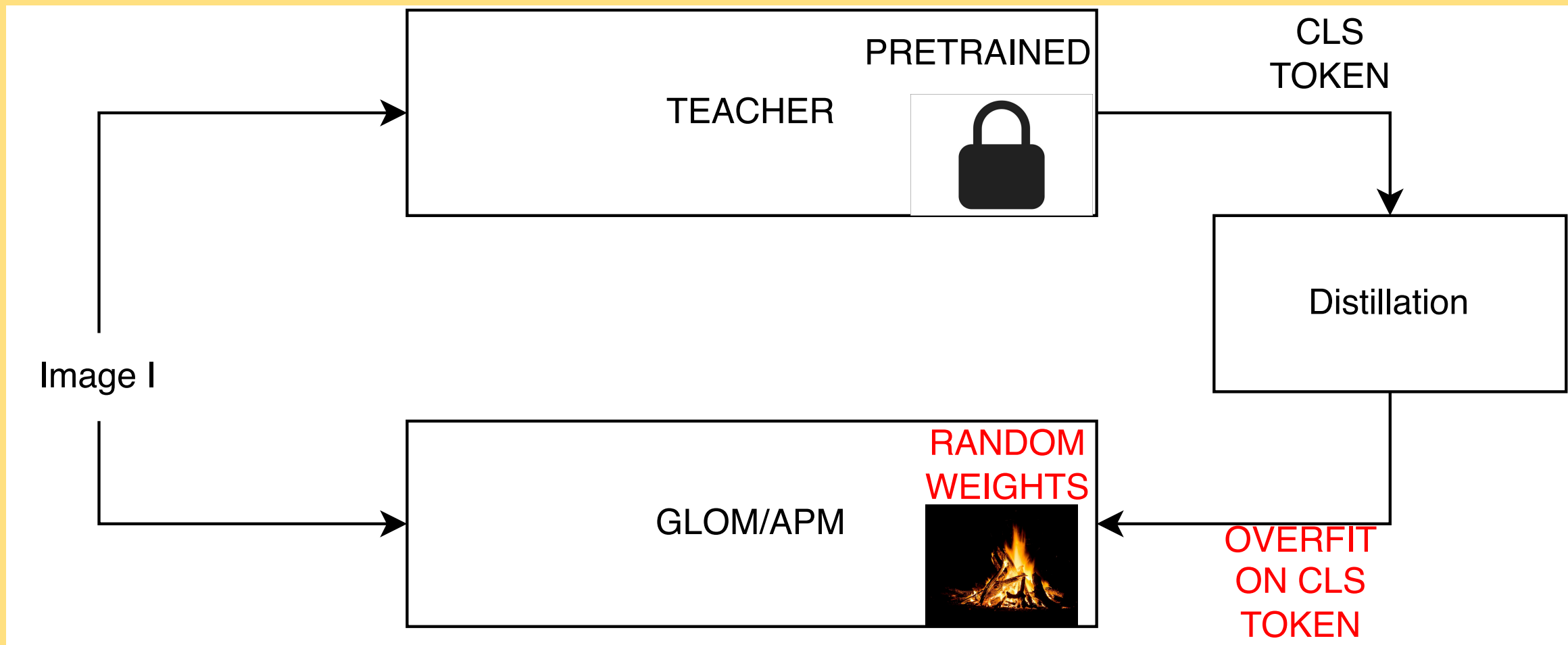
Test-Time-training

- Take a **pre-trained** model.
- Idea: there is a test sample, OOD, like corrupted with fog etc.
- Do some learning iterations on this test-sample.
- **SSL task** like rotation etc, since label cant be used.
- Classify.
- **Reset weights**
- Repeat for other test-samples.

WE do something DIFFERENT.

**- There is no other MODEL which can do
that yet.**

ONE SAMPLE-OVERFITTING



RECOVERING PATCH TOKENS FROM CLS TOKEN

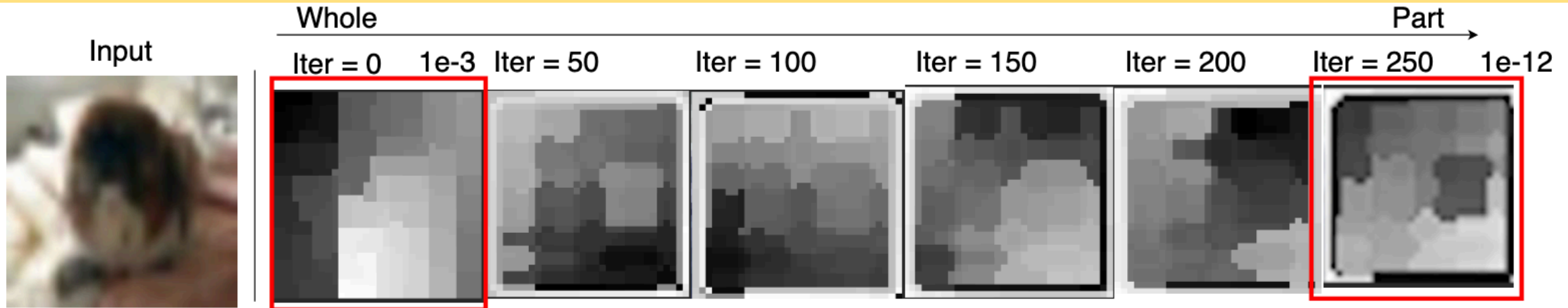
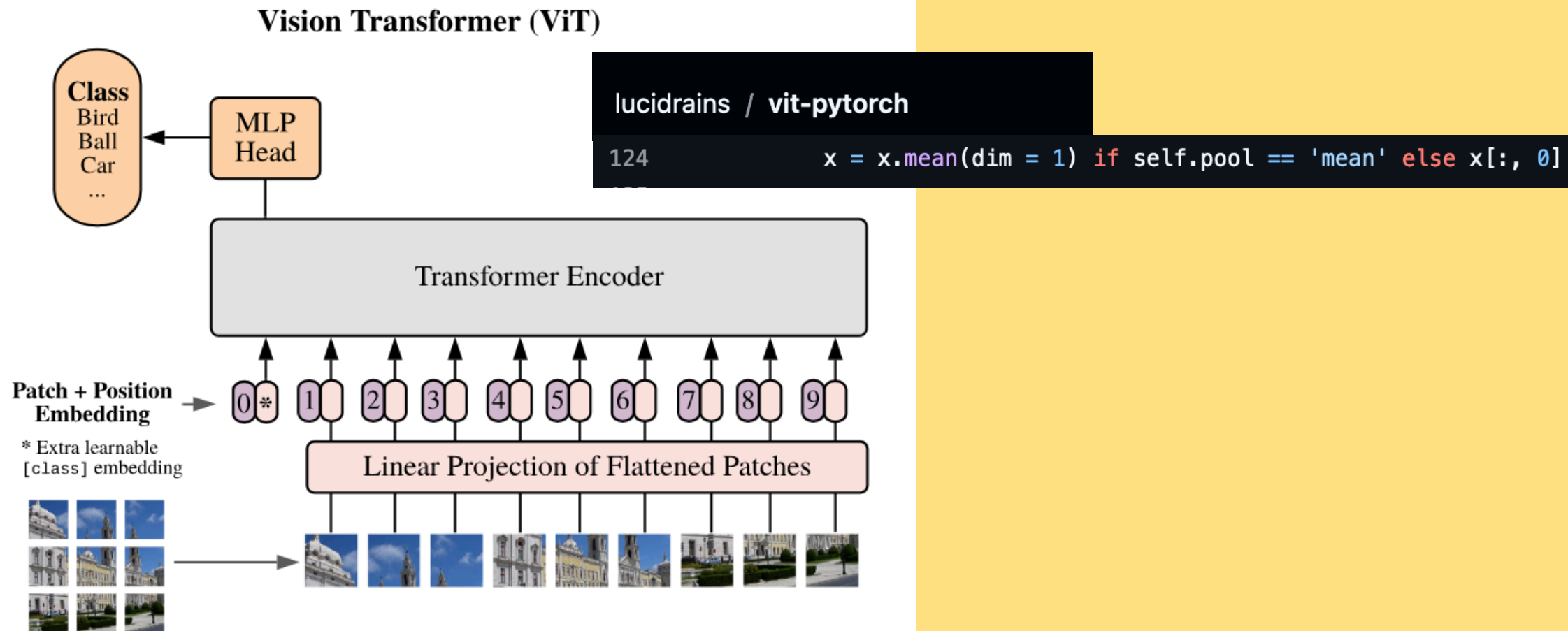


Figure 3: **Overfitting on a *single* distilled token representation leads to islands of agreement**[34]: APM is overfit on a test-sample's representation distilled from a teacher. We plot t-sne clustering of output features over 250ttt iterations. L_2 loss between predicted features and distilled sample falls from $1e-3$ to $1e-12$. Moving left to right shows that wholes break into smaller parts.

VIT DOES IT OPPOSITE.



IT SENDS INFO FROM PATCH -> CLS.

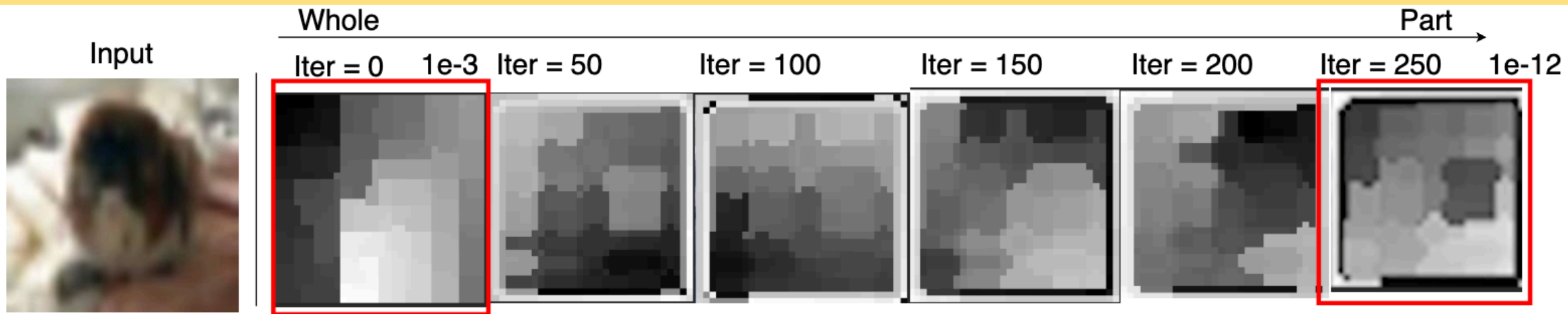
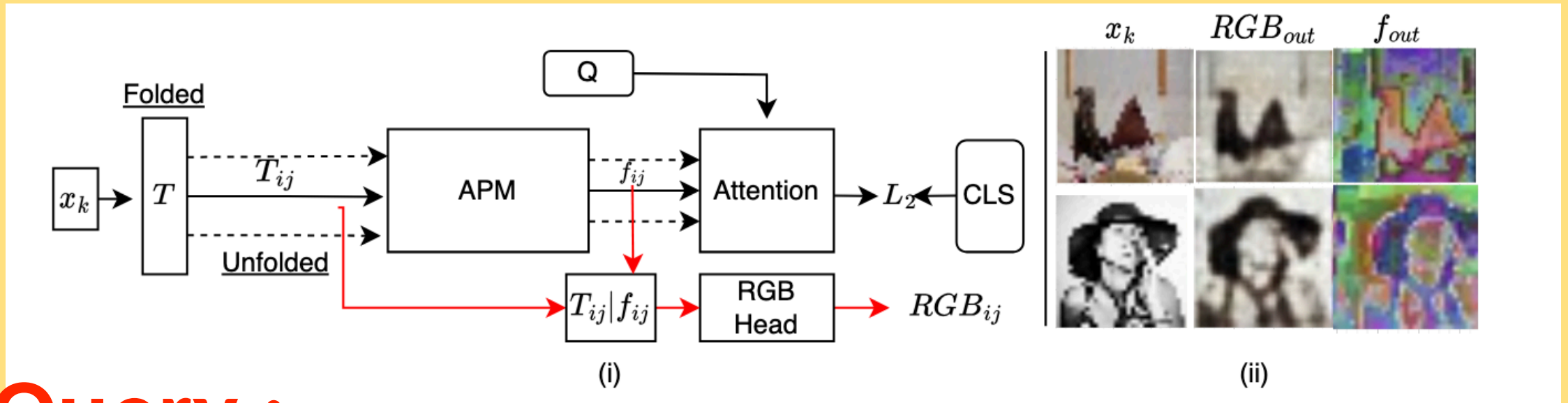


Figure 3: **Overfitting on a *single* distilled token representation leads to islands of agreement[34]:** APM is overfit on a test-sample’s representation distilled from a teacher. We plot t-sne clustering of output features over 250ttt iterations. L_2 loss between predicted features and distilled sample falls from $1e-3$ to $1e-12$. Moving left to right shows that wholes break into smaller parts.

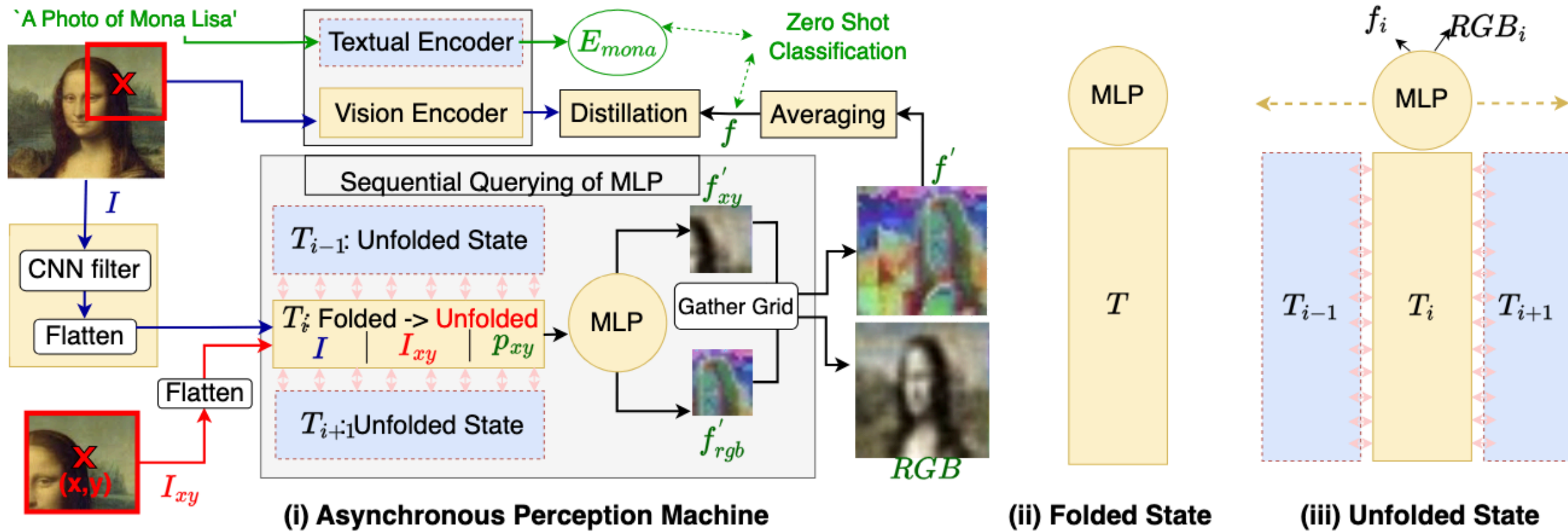
Building Object Queries At the top



1 Query :

- What is the **weight** on each predicted feature so that it **explains** the CLS token **distilled** from a **pre-trained** teacher?

The Test-Time Training Architecture



- Experiments

Table 1: **APM’s Robustness to Natural Distribution Shifts.** CoOp and CoCoOp are tuned on ImageNet using 16-shot training data per category. Baseline CLIP, prompt ensemble, TPT and our APM do not require training data. A ✓ in P means that method leveraged **pre-trained weights** on clean variant of train set aka, Image-net and downstream-ttt on corrupted version.

Method	P	ImageNet Top1 acc. ↑	ImageNet-A Top1 acc. ↑	ImageNet-V2 Top1 acc. ↑	ImageNet-R Top1 acc. ↑	ImageNet-Sketch Top1 acc. ↑	Average	OOD Average
CLIP-ViT-B/16	✗	66.7	47.8	60.8	73.9	46.0	59.1	57.2
Ensemble	✗	68.3	49.8	61.8	77.6	48.2	61.2	59.4
TPT	✗	68.9	54.7	63.4	77.0	47.9	62.4	60.8
APM (Ours)	✗	68.1	52.1	67.2	76.5	49.3	62.6	61.2
CoOp	✓	71.5	49.7	64.2	75.2	47.9	61.7	59.2
CoCoOp	✓	71.0	50.6	64.0	76.1	48.7	62.1	59.9
TPT + CoOp	✓	73.6	57.9	66.8	77.2	49.2	64.9	62.8
TPT + CoCoOp	✓	71.0	58.4	64.8	78.6	48.4	64.3	62.6
CLIP ViT-L/14	✗	76.2	69.6	72.1	85.9	58.8	72.5	71.6
APM (Ours)	✗	77.3	71.8	72.8	87.1	62.2	74.2	73.4
OpenCLIP-ViT-H/14	✗	81.6	79.1	80.7	92.9	72.8	81.4	81.3
APM (Ours)	✗	84.6	84.2	83.9	94.9	77.1	84.9	85.0

Experiments

Imagenet-C

Table 2: **APM’s performance on ImageNet-C, level 5.** The first three rows are fixed models without test-time training. The third row, ViT probing, is the baseline used in [17]. A ✓ in P means that method leveraged **pre-trained weights** on clean variant of train set aka, Image-net and downstream-ttt on corrupted version. CLIP ViT-L/14 is generally more robust. APM does better on 11/15 noises with an average accuracy score of 50.3.

	P	brigh	cont	defoc	elast	fog	frost	gauss	glass	impul	jpeg	motn	pixel	shot	snow	zoom	Average
Joint Train	✓	62.3	4.5	26.7	39.9	25.7	30.0	5.8	16.3	5.8	45.3	30.9	45.9	7.1	25.1	31.8	24.8
Fine-Tune	✓	67.5	7.8	33.9	32.4	36.4	38.2	22.0	15.7	23.9	51.2	37.4	51.9	23.7	37.6	37.1	33.7
ViT Probe	✓	68.3	6.4	24.2	31.6	38.6	38.4	17.4	18.4	18.2	51.2	32.2	49.7	18.2	35.9	32.2	29.2
TTT-MAE	✓	69.1	9.8	34.4	50.7	44.7	50.7	30.5	36.9	32.4	63.0	41.9	63.0	33.0	42.8	45.9	44.4
OpenCLIP ViT-L/14	✗	71.9	47.0	50.3	32.7	58.3	46.9	26.0	26.5	28.1	62.7	37.7	58.3	28.2	50.4	37.9	42.1
APM (Ours)	✗	77.4	51.9	56.6	37.9	64.8	53.2	28.7	31.4	33.0	68.4	44.1	64.5	33.1	56.9	43.9	50.3

Experiments

Cross-Dataset Generalization

Table 3: **Cross-dataset generalization** from ImageNet to fine-grained classification datasets. CoOp and CoCoOp are tuned on ImageNet using 16-shot training data per category. Baseline CLIP, prompt ensemble, TPT and APM do not require training data or annotations. We report top-1 accuracy.

Method	P	Flower102	DTD	Pets	UCF101	Caltech101	Food101	SUN397	Aircraft	EuroSAT	Average
CoOp	✓	68.7	41.9	89.1	66.5	93.7	85.3	64.2	18.5	46.4	63.9
CoCoOp	✓	70.9	45.5	90.5	68.4	93.8	84.0	66.9	22.3	39.2	64.6
CLIP-ViT-B/16	✗	67.4	44.3	88.3	65.1	93.4	83.7	62.6	23.7	42.0	63.6
Ensemble	✗	67.0	45.0	86.9	65.2	93.6	82.9	65.6	23.2	50.4	64.6
TPT	✗	69.0	47.8	87.8	68.0	94.2	84.7	65.5	24.8	42.4	65.1
APM (Ours)	✗	62.0	48.9	81.6	72.6	89.6	84.2	65.7	29.7	55.7	65.5

APM Feature-Analysis

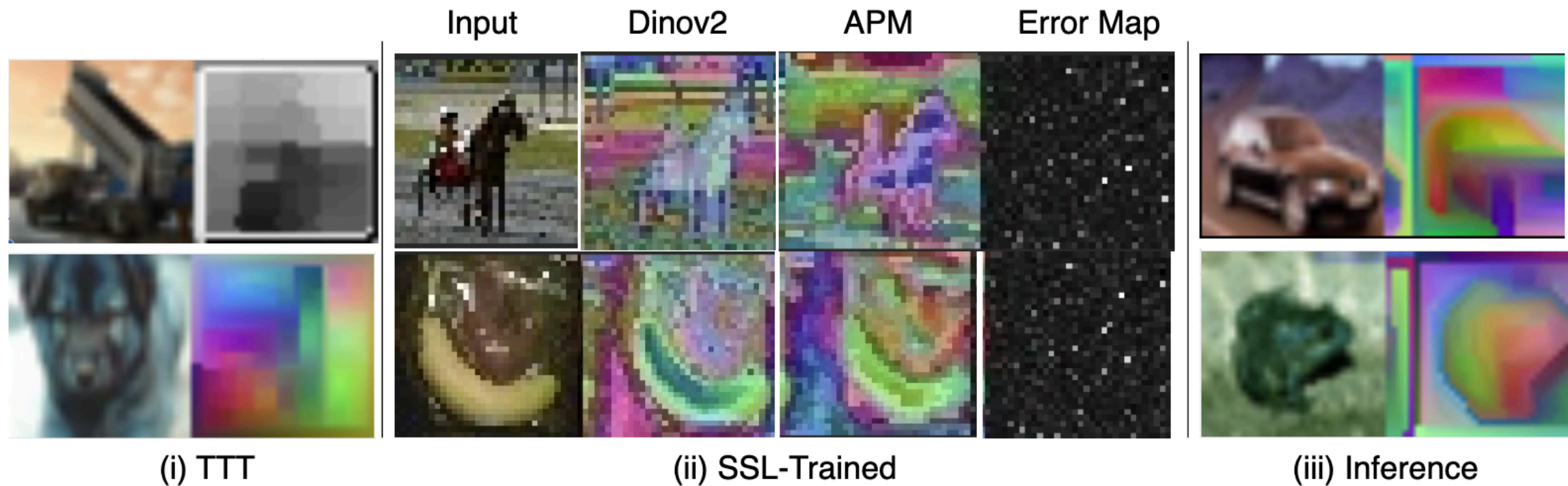
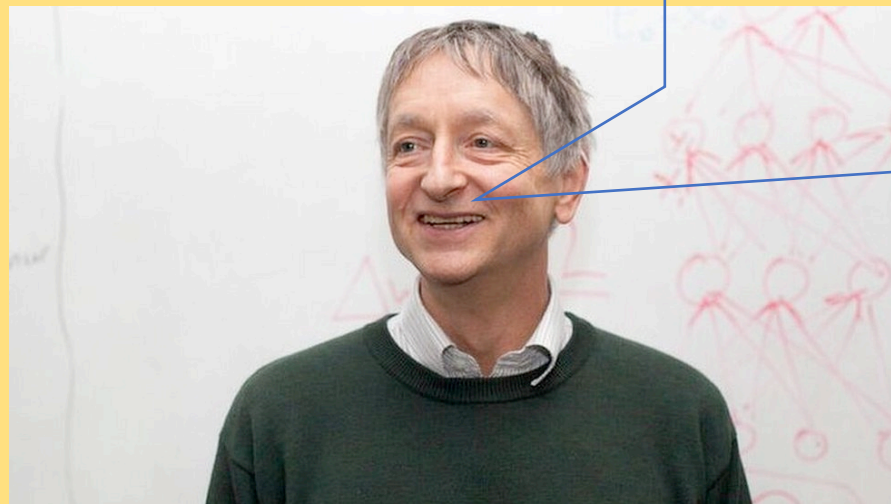
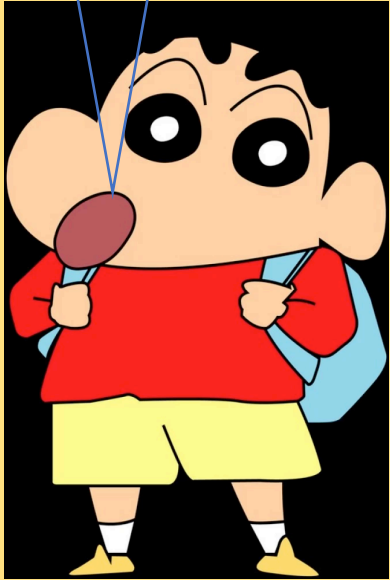


Figure 5: **APM feature Analysis:** (i) TTT iterations on an input image leads to semantically aware clustering. top: 2D t-sNE. bottom: 3D t-sNE. [70, 34]. (ii) APM is trained via self-supervision using DINOv2-Teacher. (from left) Input, Dinov2 grid, APM grid. APM's grid **closely approximates** Dinov2 grid evident from black regions in error map. Note that APM does asynchronous patch-based processing whereas Dinov2 does parallel perception. (iii) Cifar-10 samples feed-forwarded through SSL-trained APM yields features of significant semantic quality.[34]

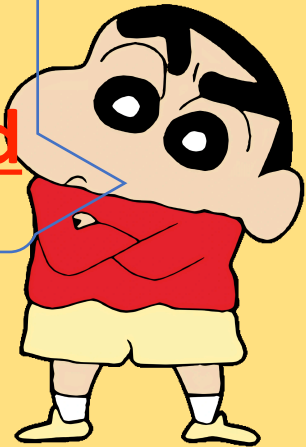


Backprop?
Eww,
Gradient Descent?
Ewwww
Eww, Ewww, Ewww,
Geoff

Geoff sir,
Happy now?
GLOM works.




Ewww? You are NOT
easily satisfied eh??
Go find some other kid




Ewww Continued.....


The Deep Learning Saga



Yoshua Bengio
549 subscribers

Subscribe

 **@rvillegass** 11 years ago
Dr. Hinton aged, but that girl never aged. AMAZING!

4  Reply

Girl who never ages.

Huntron I finally know how the brain works.



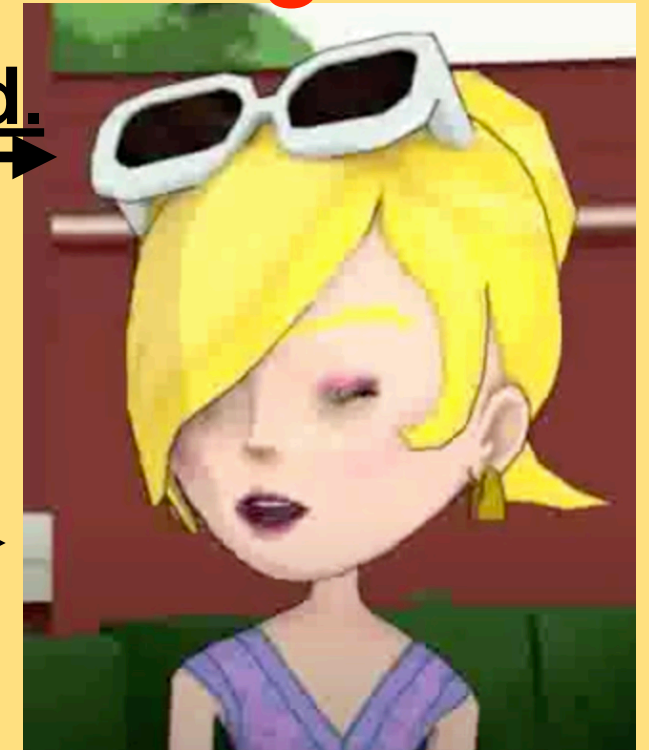
2024

It's a GLOM of all i ever said.

Whatever.

Really, i am serious this time. Pinky swear.

Ewwwwwwwwww



The one
Who **NEVER**
AGES.

Questions? 🙄

Nothing is ewwww

Weirdos like us





Poster

Asynchronous Perception Machine for Test Time Training

Rajat Modi · Yogesh Rawat

East Exhibit Hall A-C #2103

[[Abstract](#)] [[Project Page](#)]

Wed 11 Dec 4:30 p.m. PST – 7:30 p.m. PST ([Bookmark](#))