CapsNets Continuing the Convolutional Quest

Sascha Diefenbacher, Herman Frost, Gregor Kasiezcka,
Tilman Plehn, Jennifer M. Thompson

Based on: 1906.11265

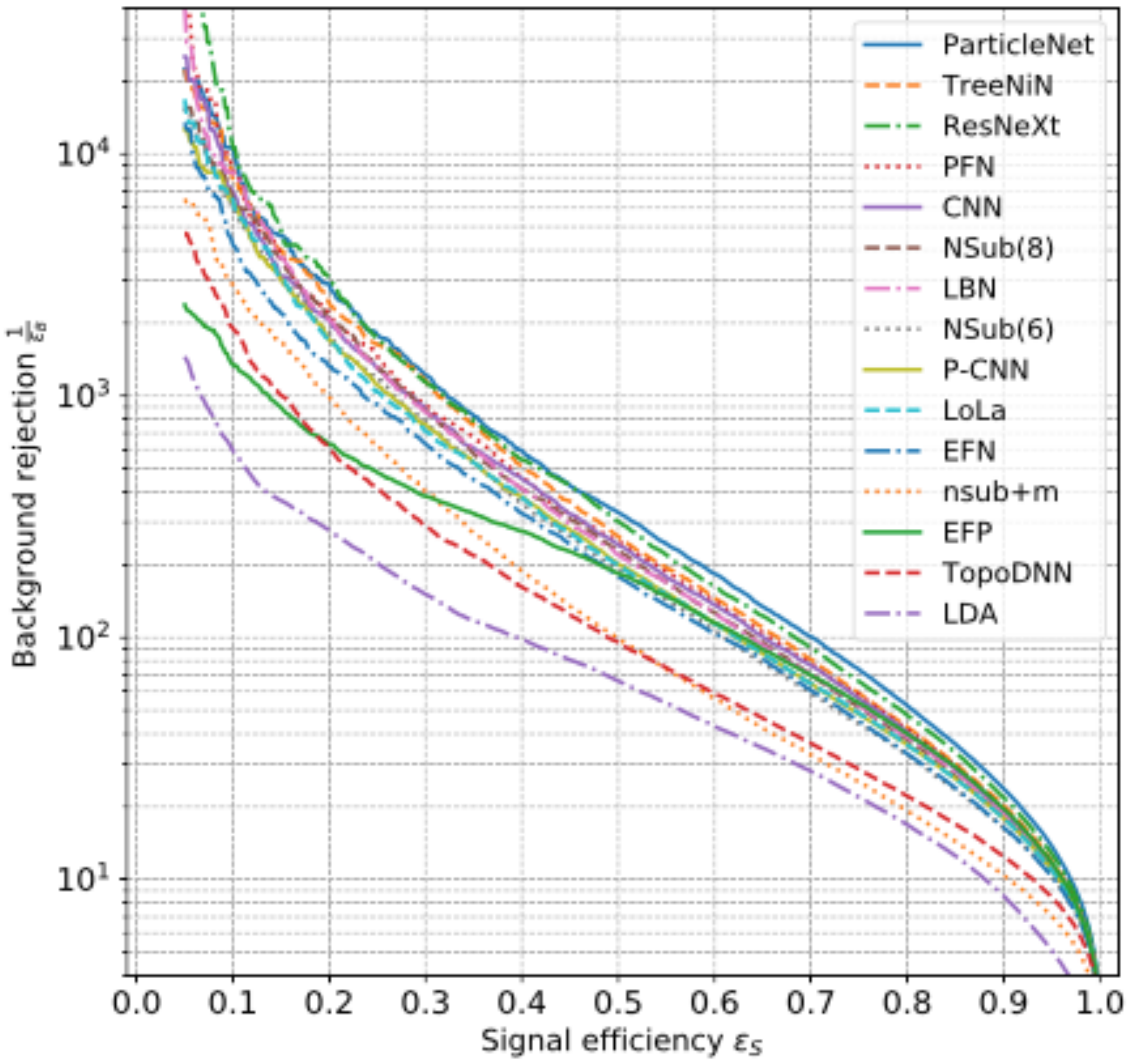
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

DESY.

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

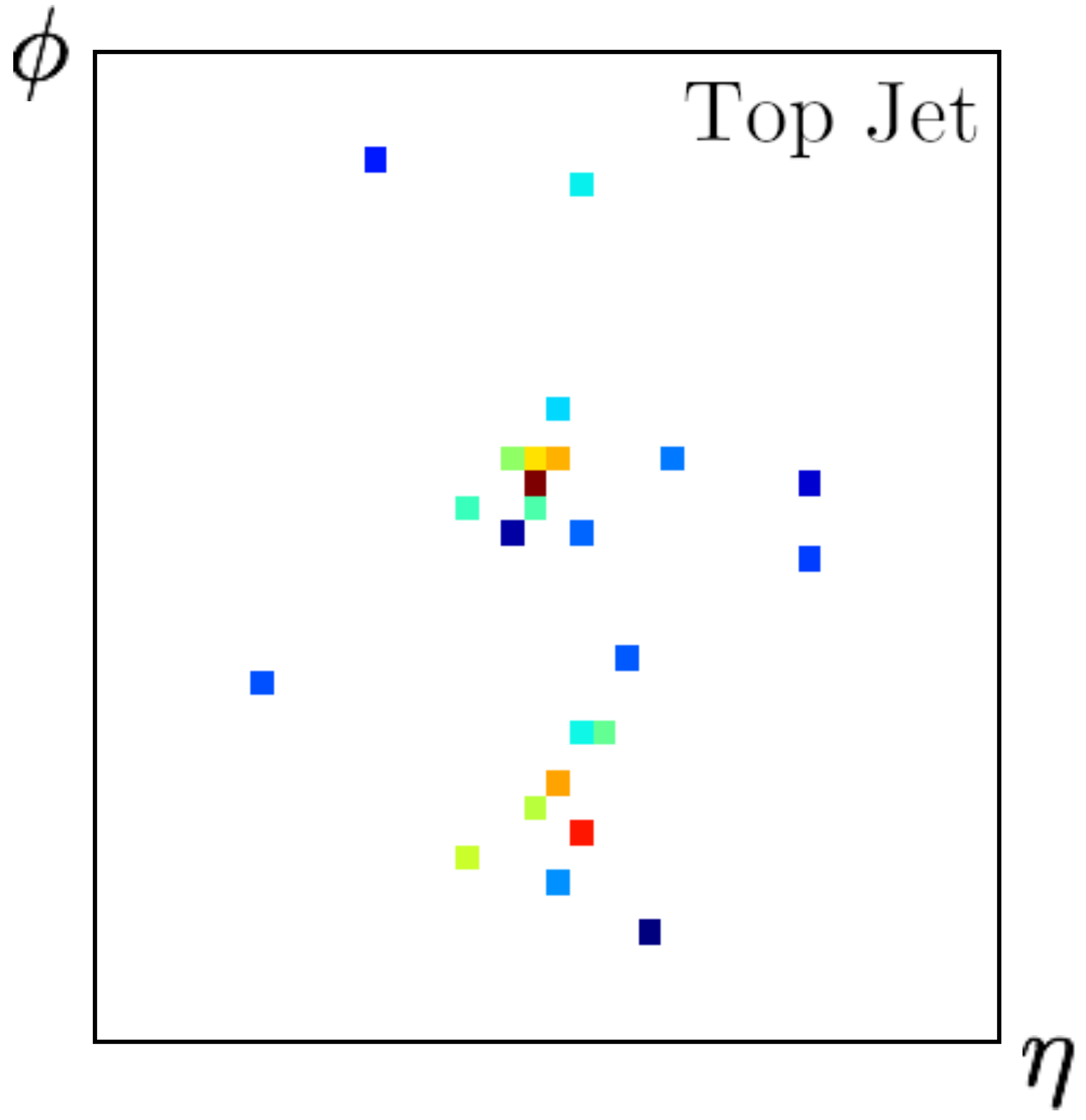
Beyond Performance

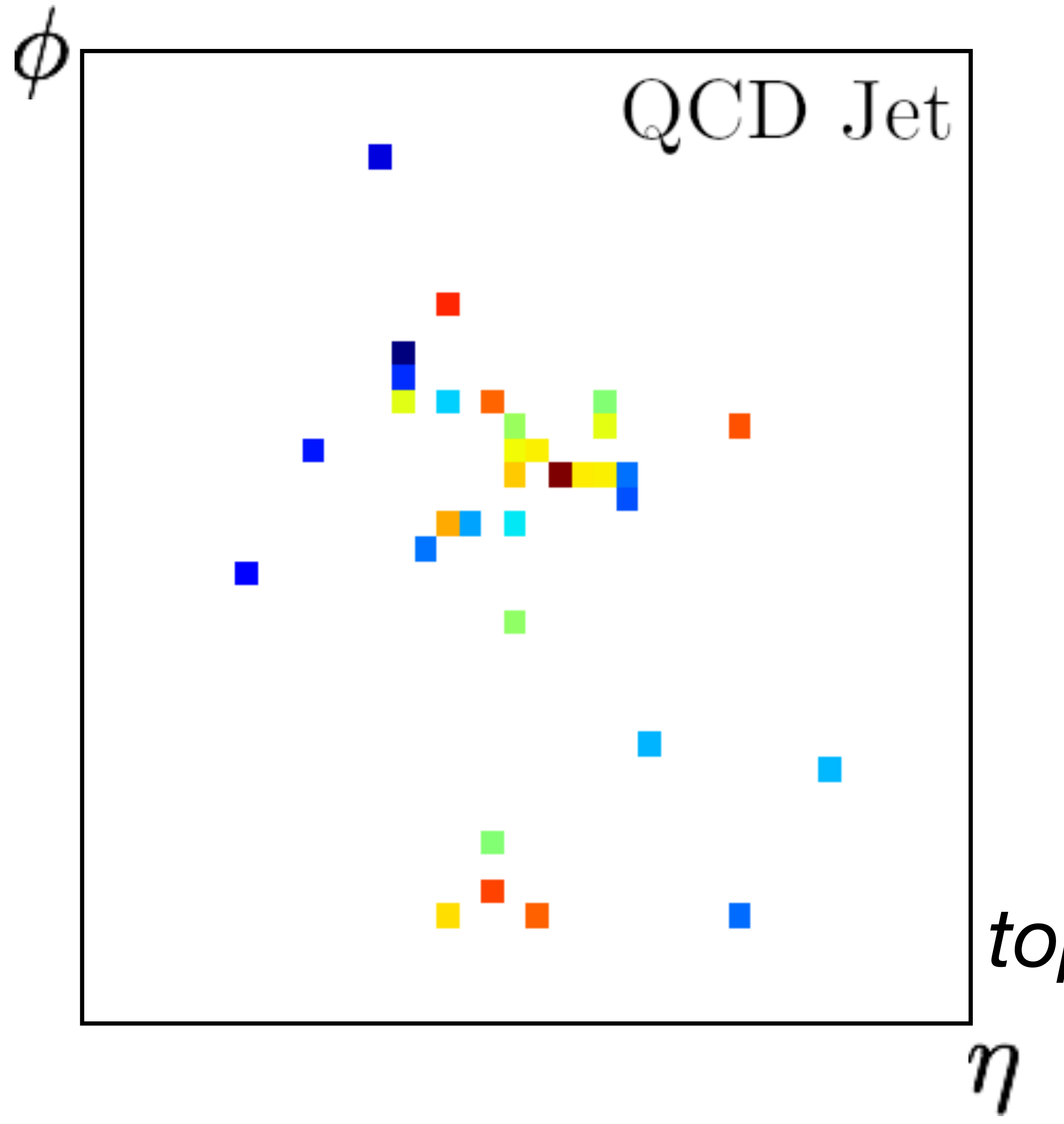


1902.09914

- Neural network classifiers proven and invaluable tool in particle physics
- Movement towards calibration, stability and insight

Beyond Performance



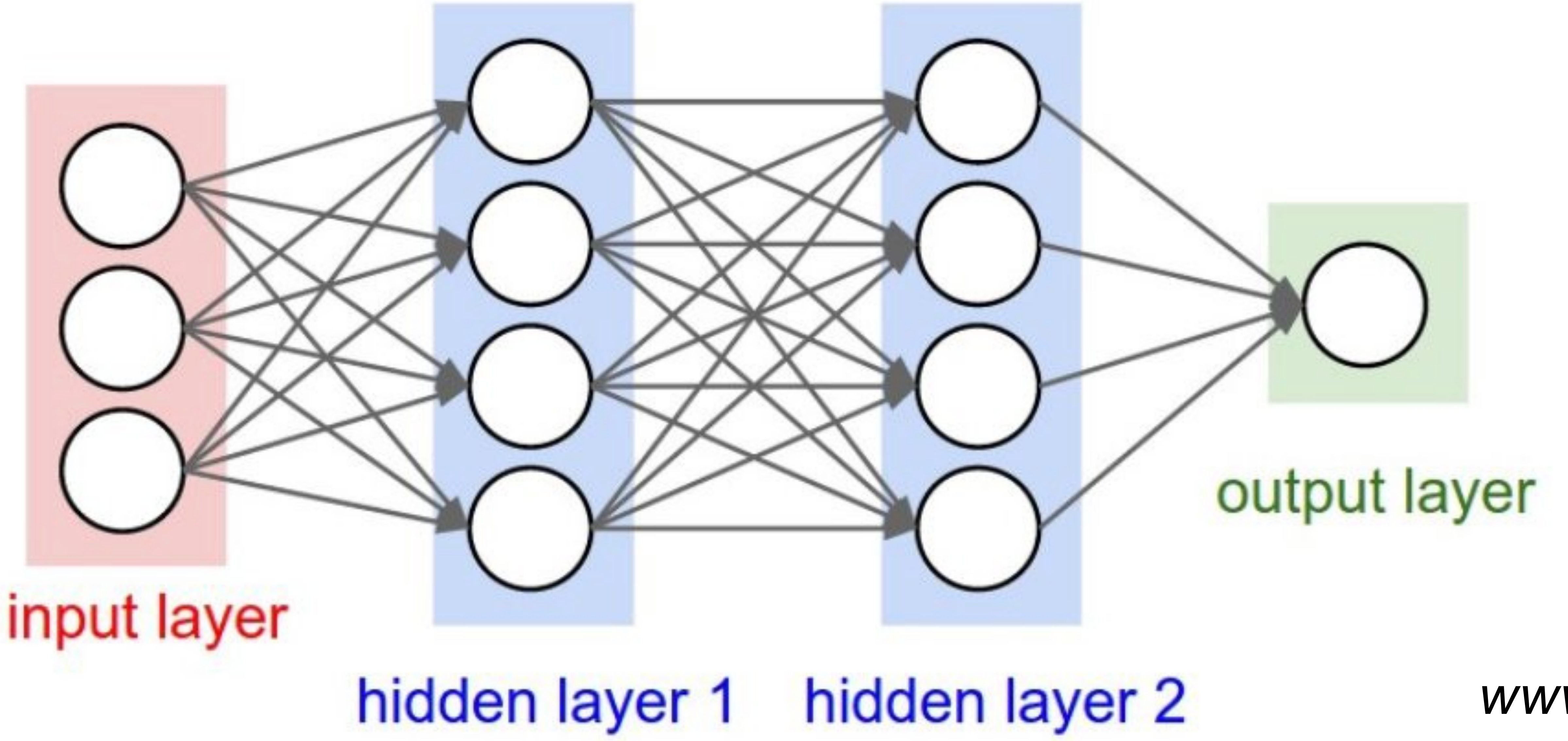


top tagging reference data

- Understanding network decision process

➡ Insight into underlying physics

Understanding Decisions



www.digitaltrends.com

- Standard DNN: Only value of the output has meaning
- Intermediate not individually expressive

➔ Difficult to interpret

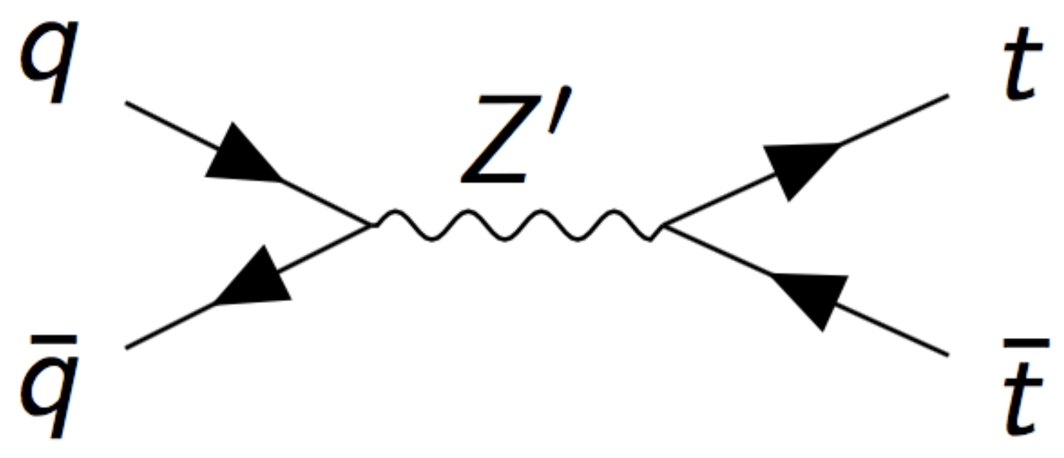
➔ New approach: Capsule Networks

Full Event Tagging

- Capsules well suited for full events
- Event level kinematics easier to understand than jet variables

Signal:

Z' (1 TeV) decaying to top pairs



Background:

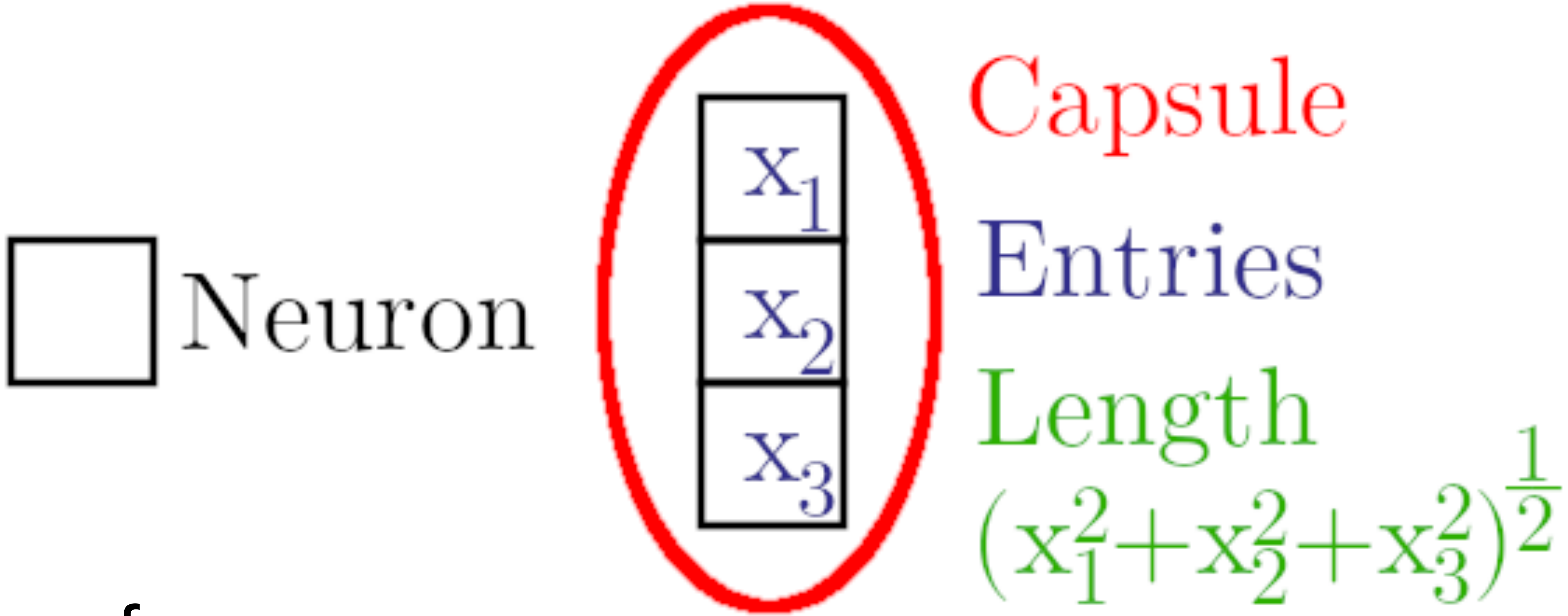
QCD di-top processes

$$pp \rightarrow t\bar{t}\ [QCD]$$

- Preselection:
  - Number of jets $\geqslant$ 2 $(\mathrm{anti}\ k_t, \Delta R = 1.0)$
  - $p_{T jet1,2} >$ 350 GeV
  - $|\eta_{jet1,2}| <$ 2.0

Capsule Networks

Capsule Networks



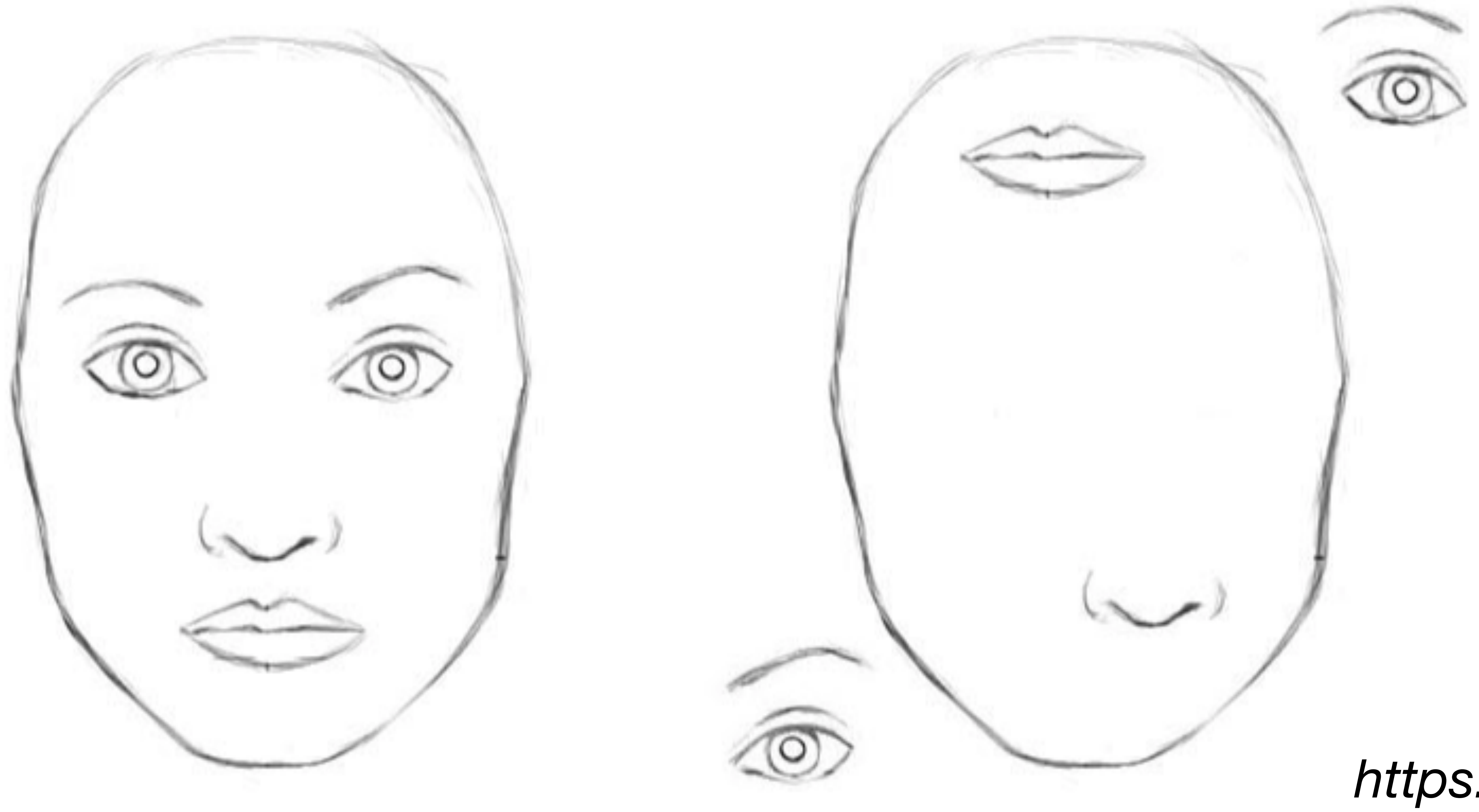
- Capsule: Group of neurons
- Capsule outputs describe instantiation vectors
  - Entries of vector describes properties of object
  - Length corresponds to probability of object existence

➡ Interpretability through meaningful entries

S. Sabour, N. Frosst, G. E. Hinton: **Dynamic Routing Between Capsules:** 1710.09829

Further Features

More information than just feature presence

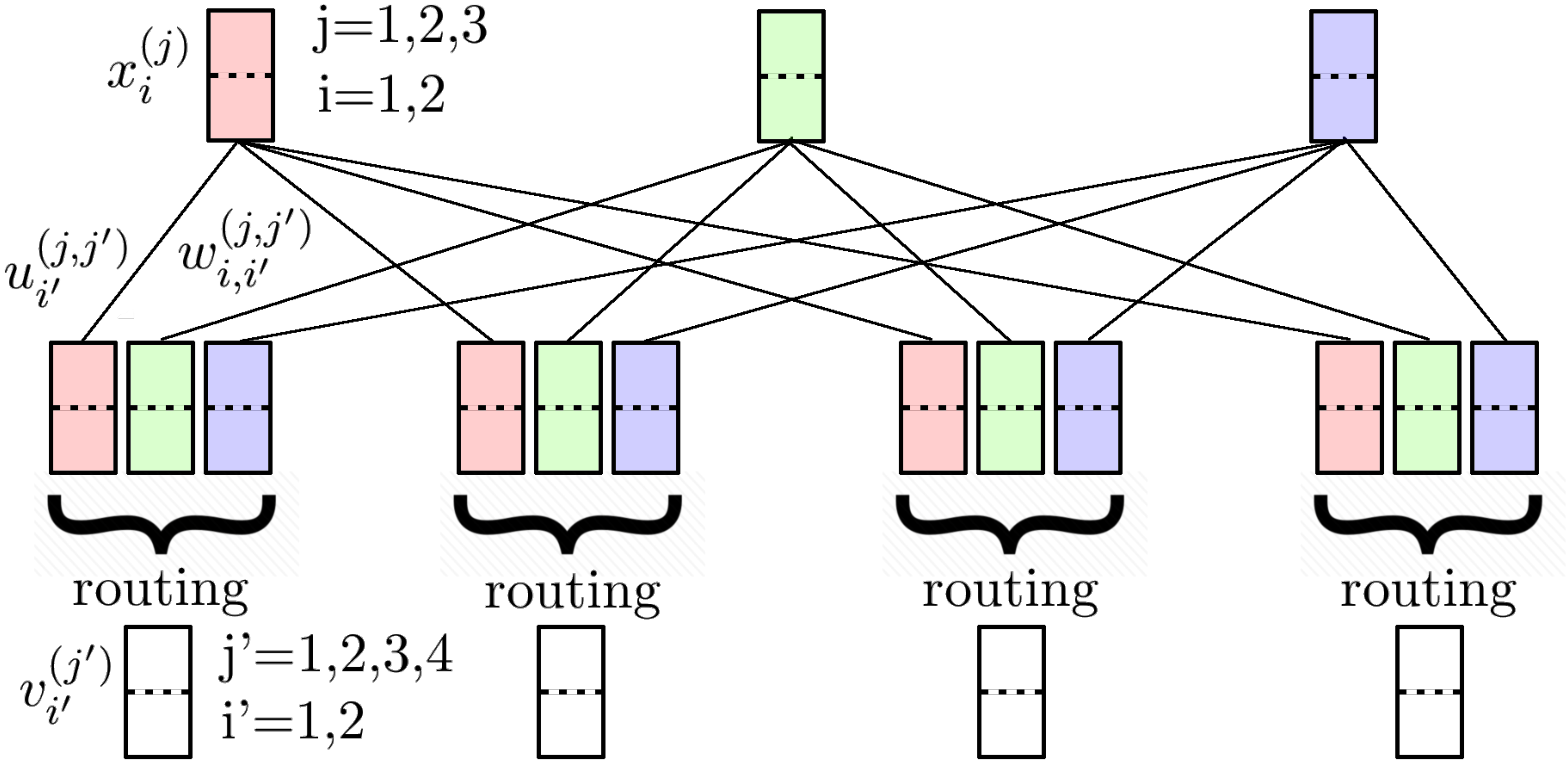


https://pechyonkin.me/capsules-1/

➡ Relative position of features taken into account

➡ Can learn event level kinematics

Capsule Layers

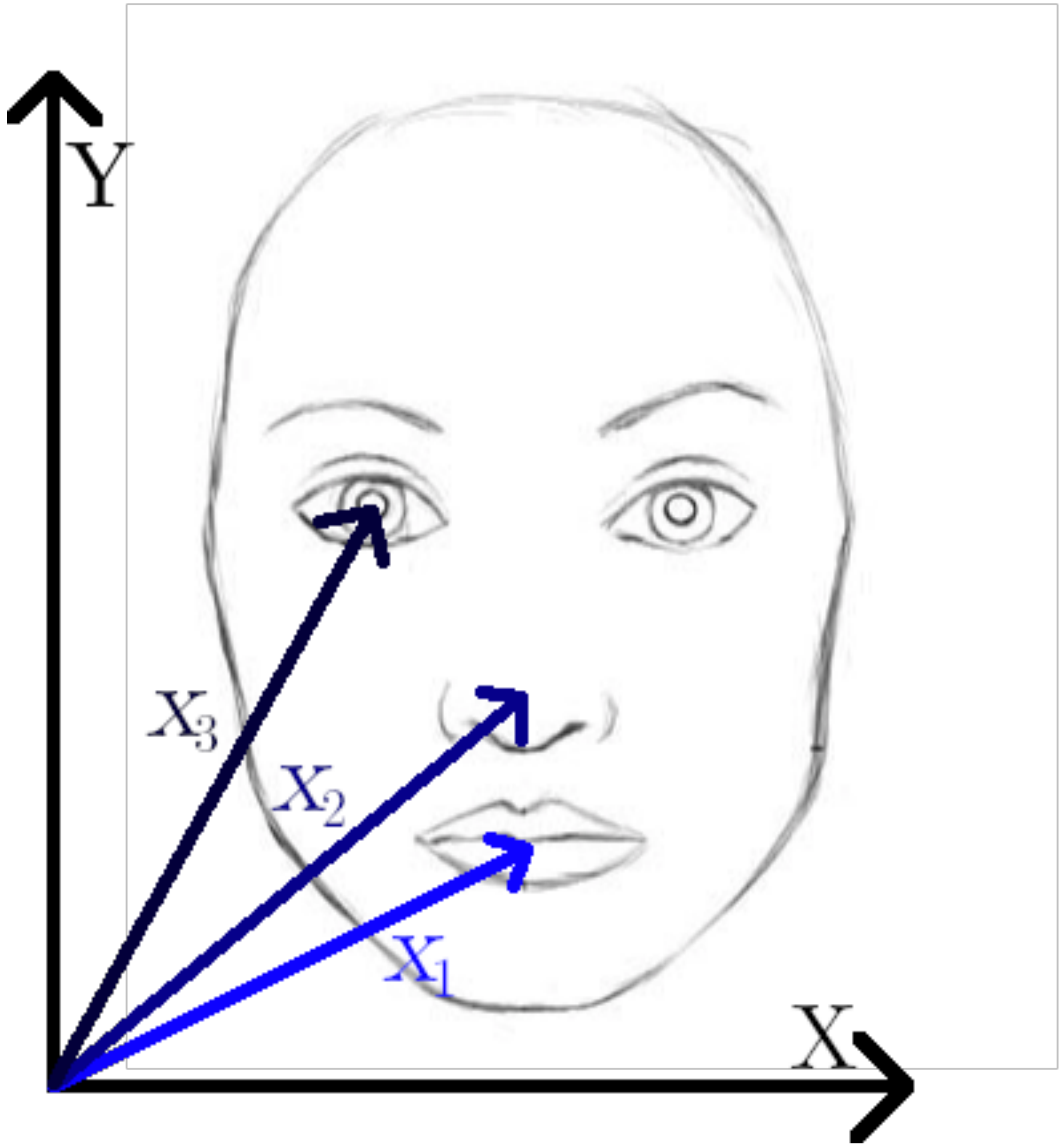


- Each wector weighted by matrix $w_{i,i'}$
- Contributions combined through dynamic routing

Routing Intuition

Example: Human Face

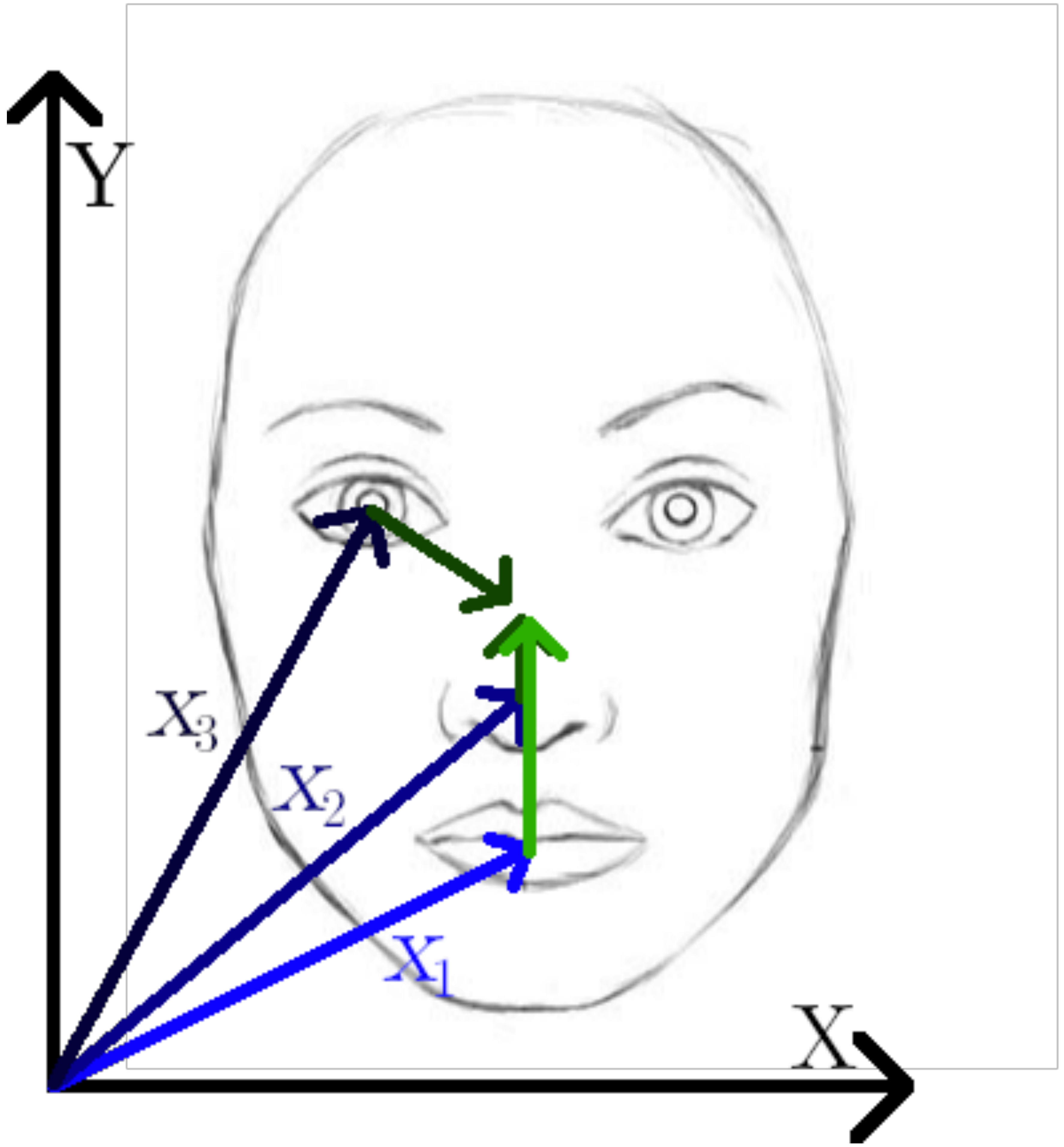
- 3 Input capsules describe
  1. Mouth
  2. Nose
  3. Left eye
- Entires: X and Y positions
- Aim: Combine into face capsule



Routing Intuition

Example: Human Face

- Capsules make prediction about face position



Routing Intuition

Example: Human Face

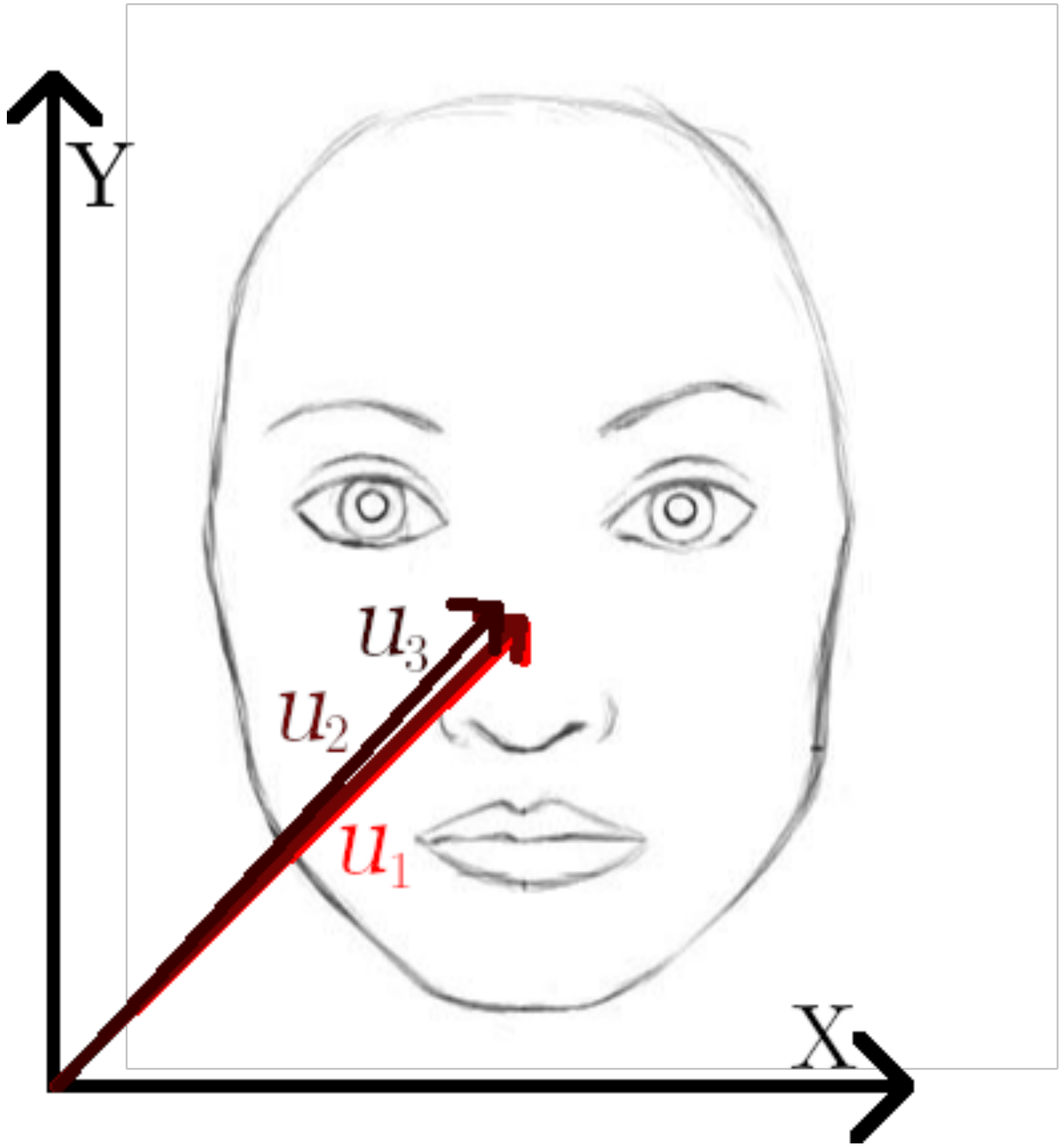
- Capsules make prediction about face position
- Predicted positions given by Capsule times weight matrix

$$u_j = \mathrm{W}_{j,1} x_j$$

- ➡ Predictions agree

➡ Combined face capsule large

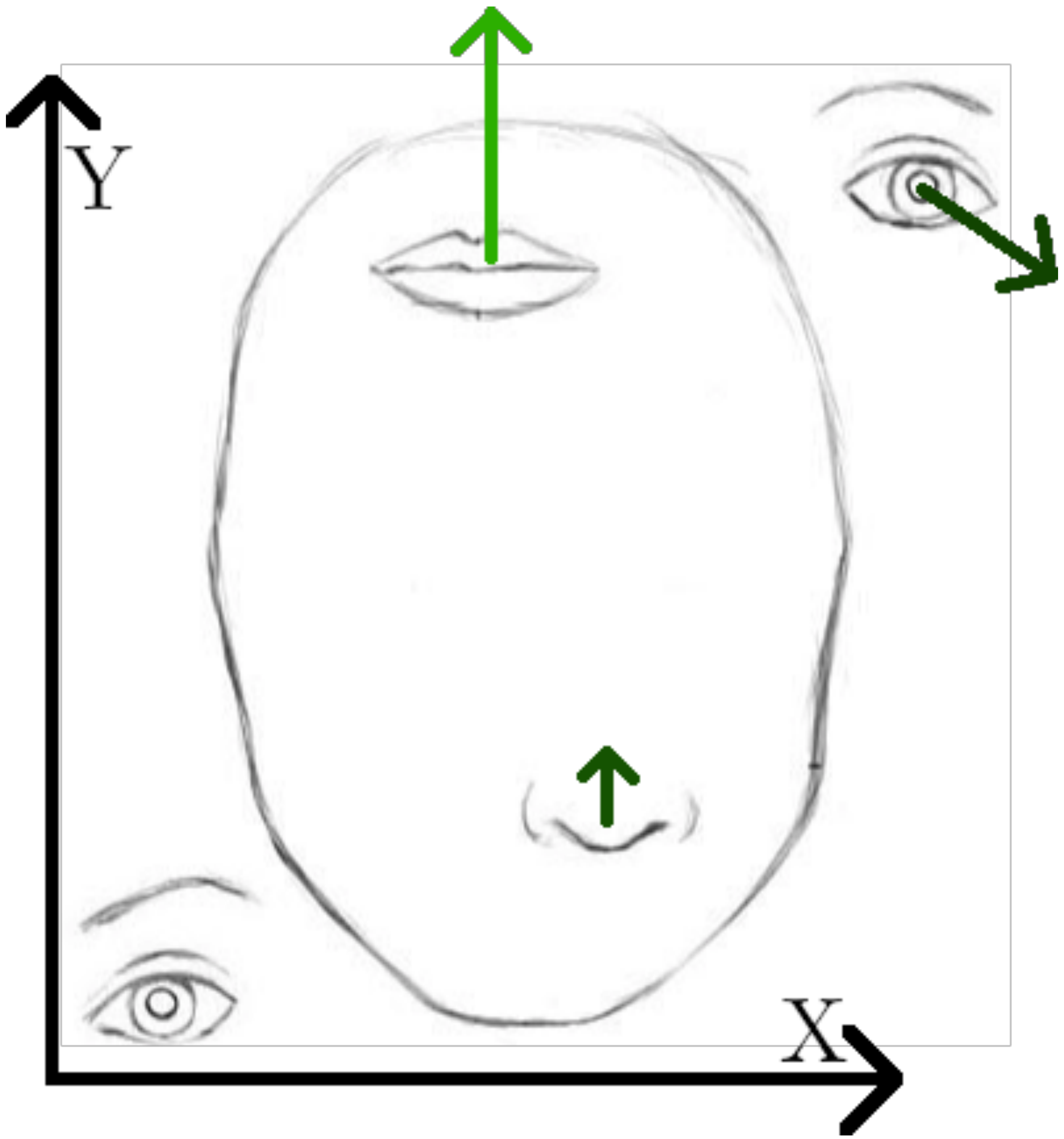
Classified as face



Routing Intuition

Example: Human Face

- Predictions are off if features at wrong relative positions



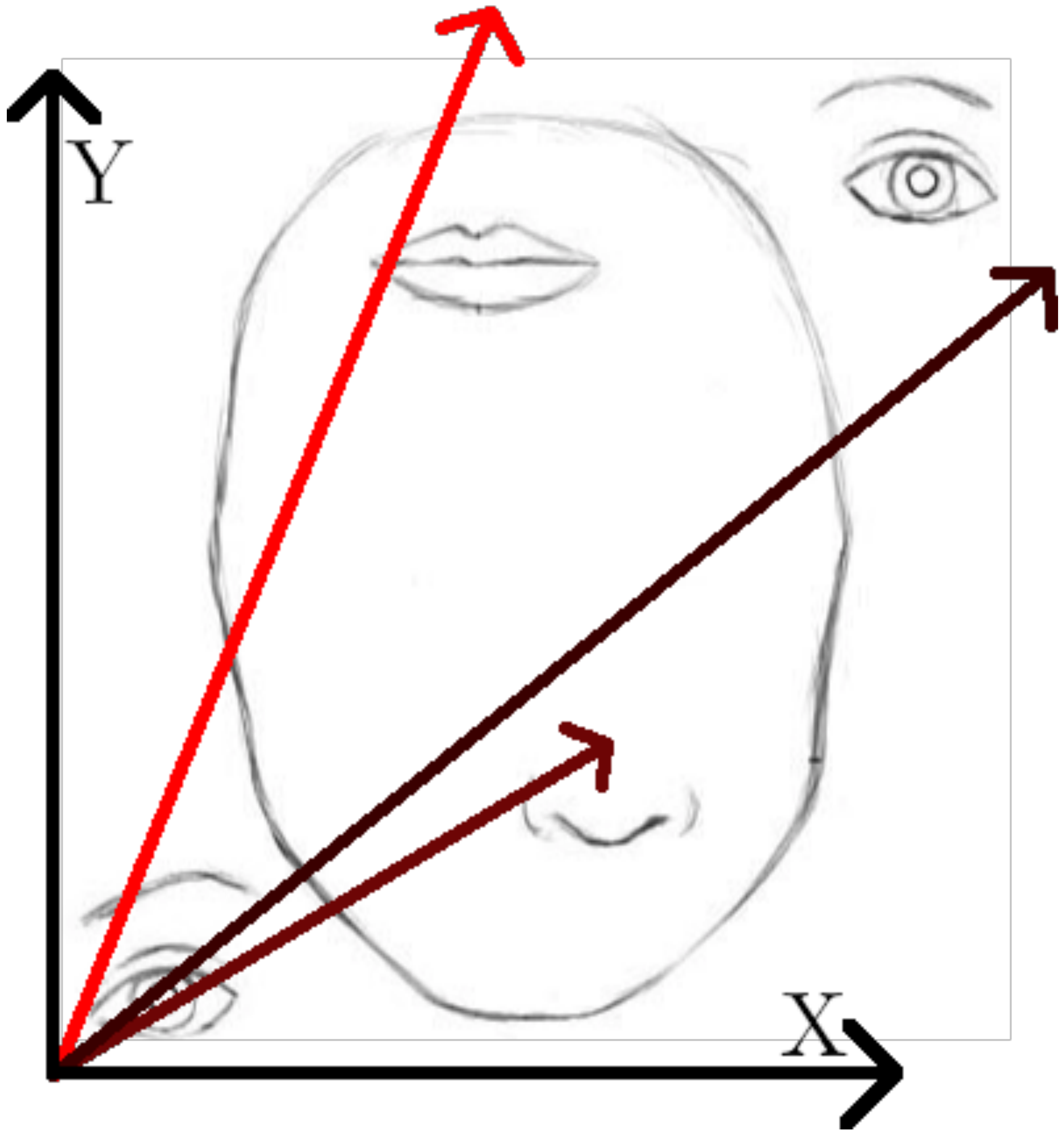
Routing Intuition

Example: Human Face

- Predictions are off if features at wrong relative positions
- Predicted positions do not agree

➔Face capsule should short

➔Classified as not-face



➔Need algorithm that enhances agreeing contributions

Dynamic Routing

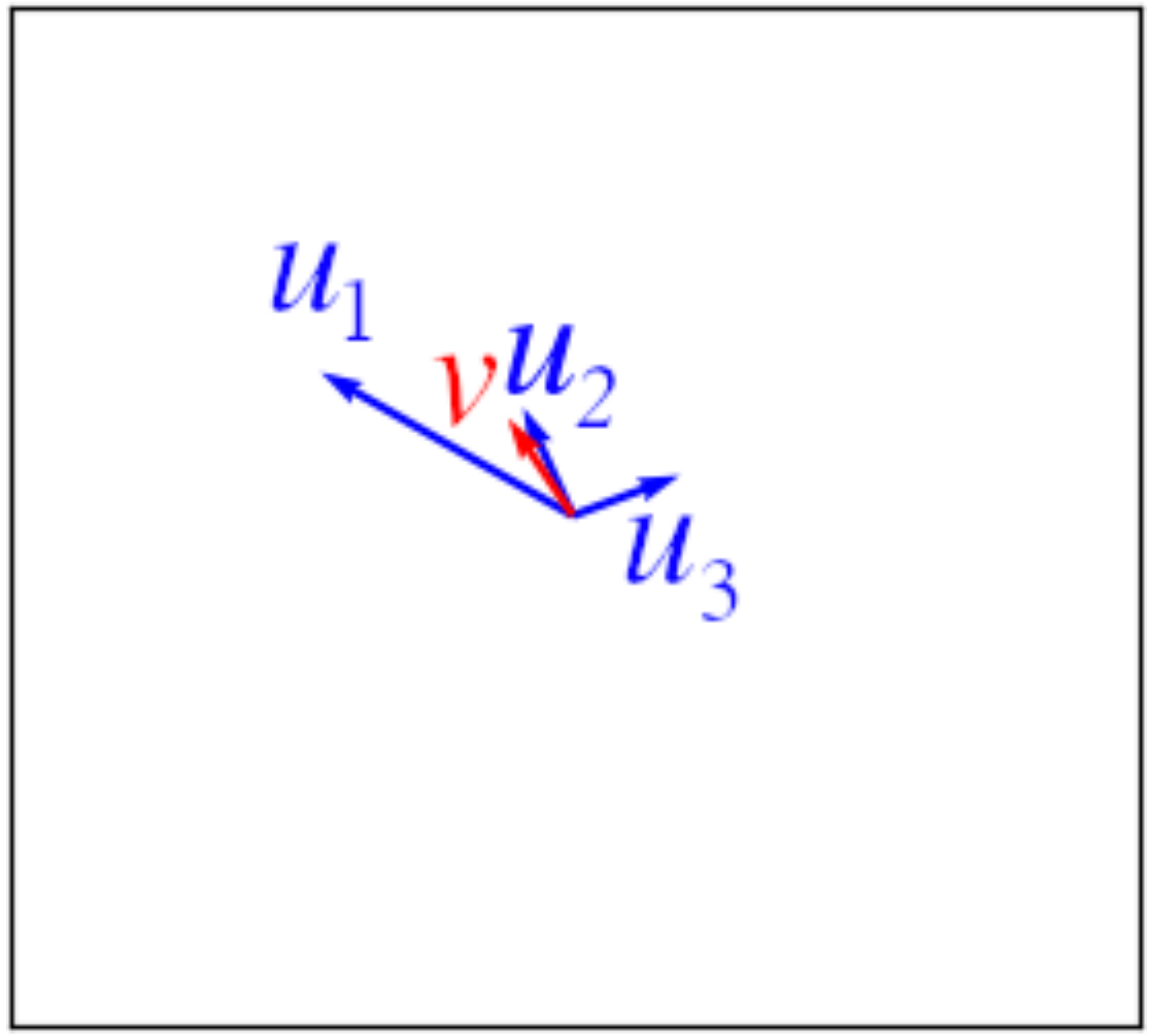
Routing by agreement algorithm:

- Calculate weighted average

$$v = \sum_j c_{(j)} u_{(j)}$$

- Update weights $c_{(j)}$ based on $v \cdot u_{(j)}$

Before Routings



Dynamic Routing

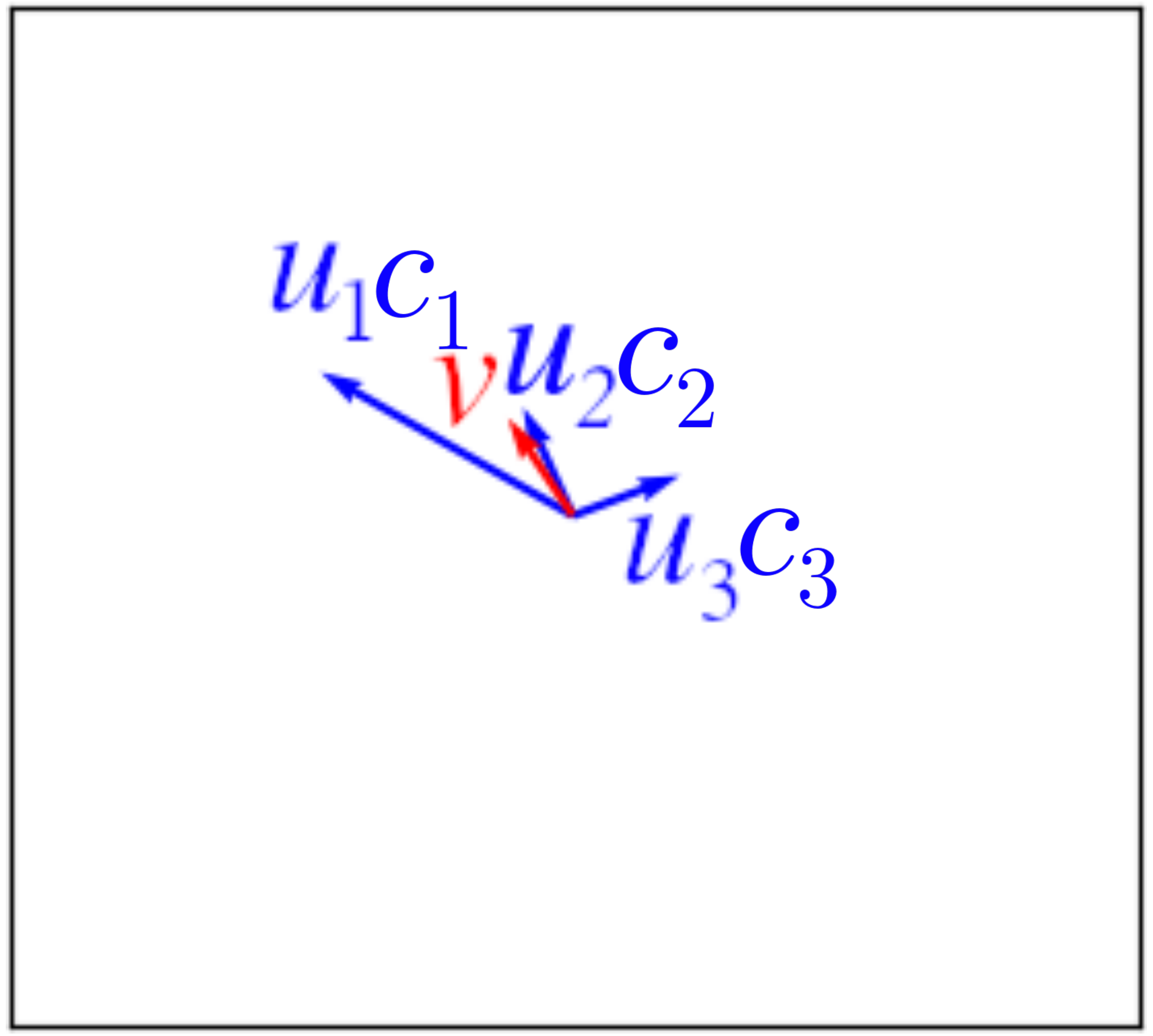
Routing by agreement algorithm:

- Calculate weighted average

$$v = \sum_j c_{(j)} u_{(j)}$$

- Update weights $c_{(j)}$ based on $v \cdot u_{(j)}$

Before Routings



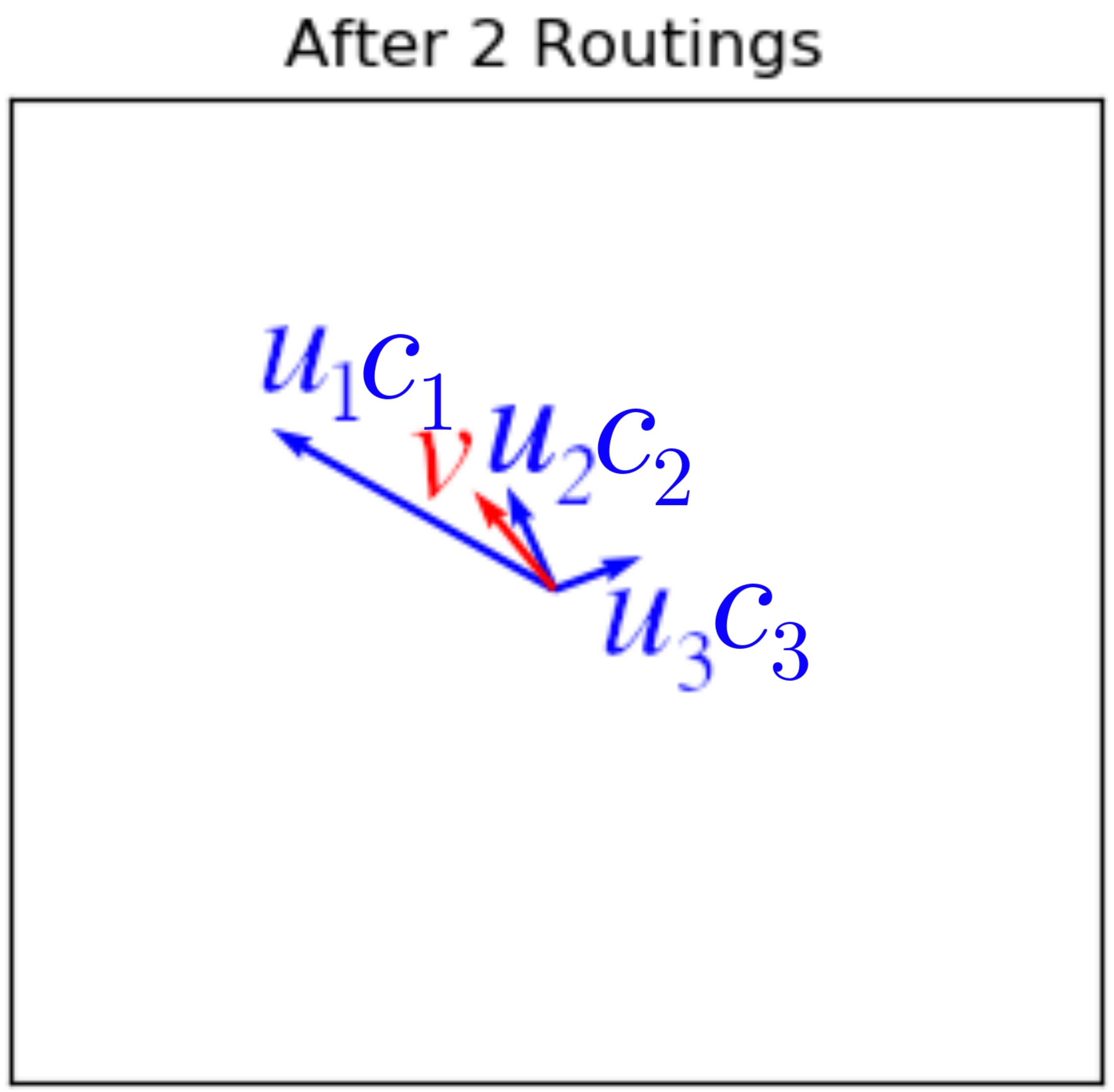
Dynamic Routing

Routing by agreement algorithm:

- Calculate weighted average

$$v = \sum_j c_{(j)} u_{(j)}$$

- Update weights $c_{(j)}$ based on $v \cdot u_{(j)}$



Dynamic Routing

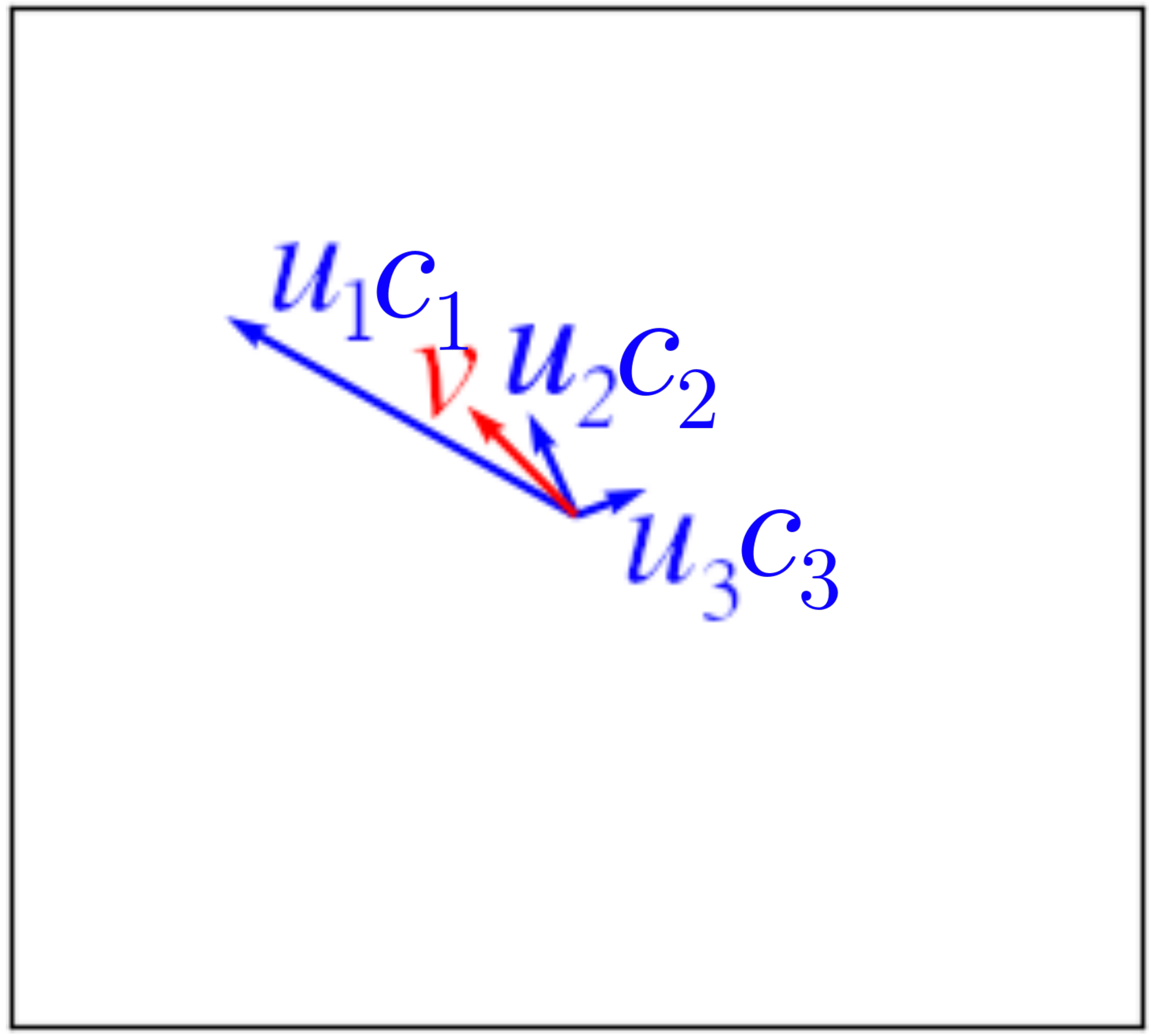
Routing by agreement algorithm:

- Calculate weighted average

$$v = \sum_j c_{(j)} u_{(j)}$$

- Update weights $c_{(j)}$ based on $v \cdot u_{(j)}$

After 4 Routings



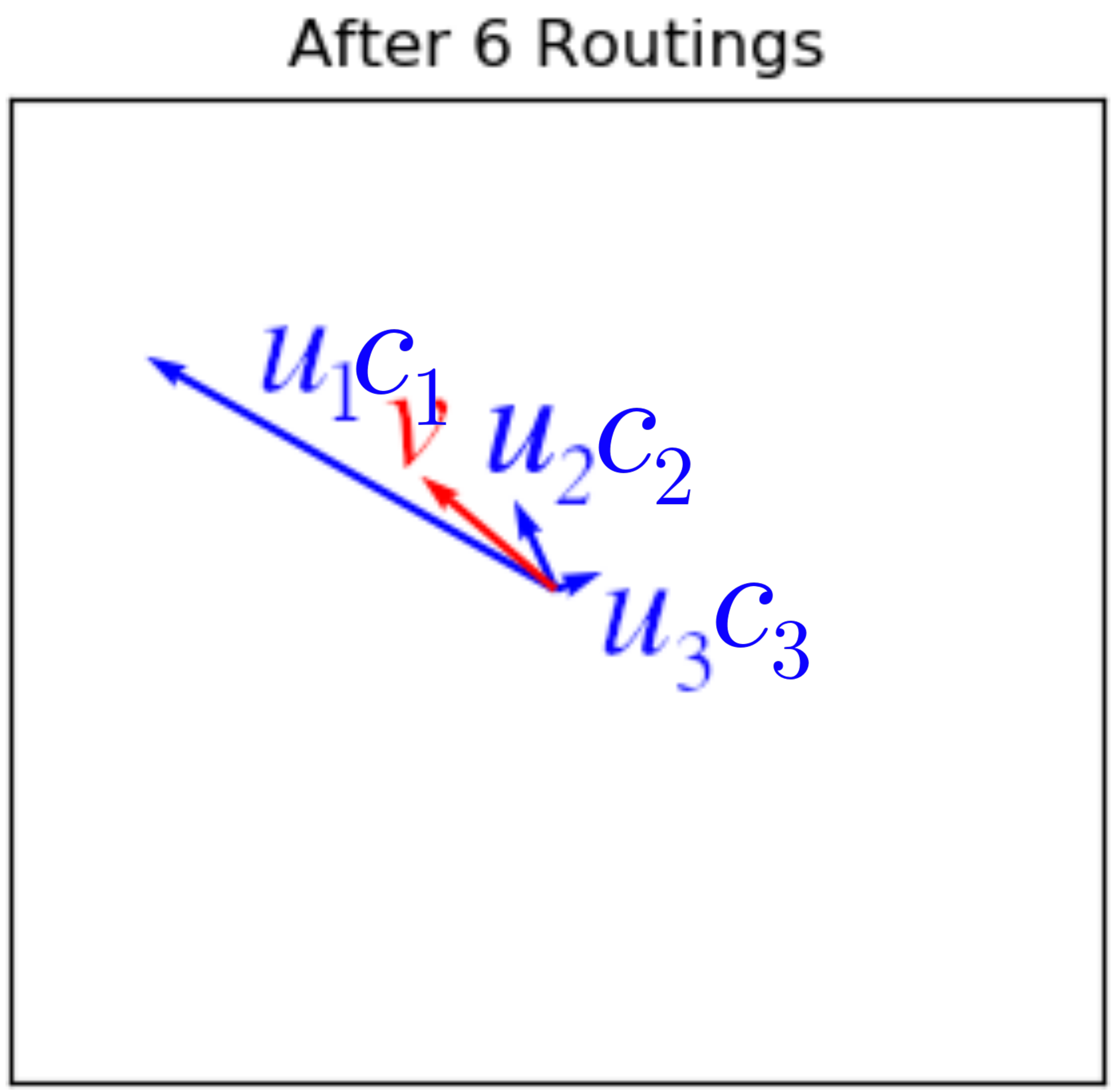
Dynamic Routing

Routing by agreement algorithm:

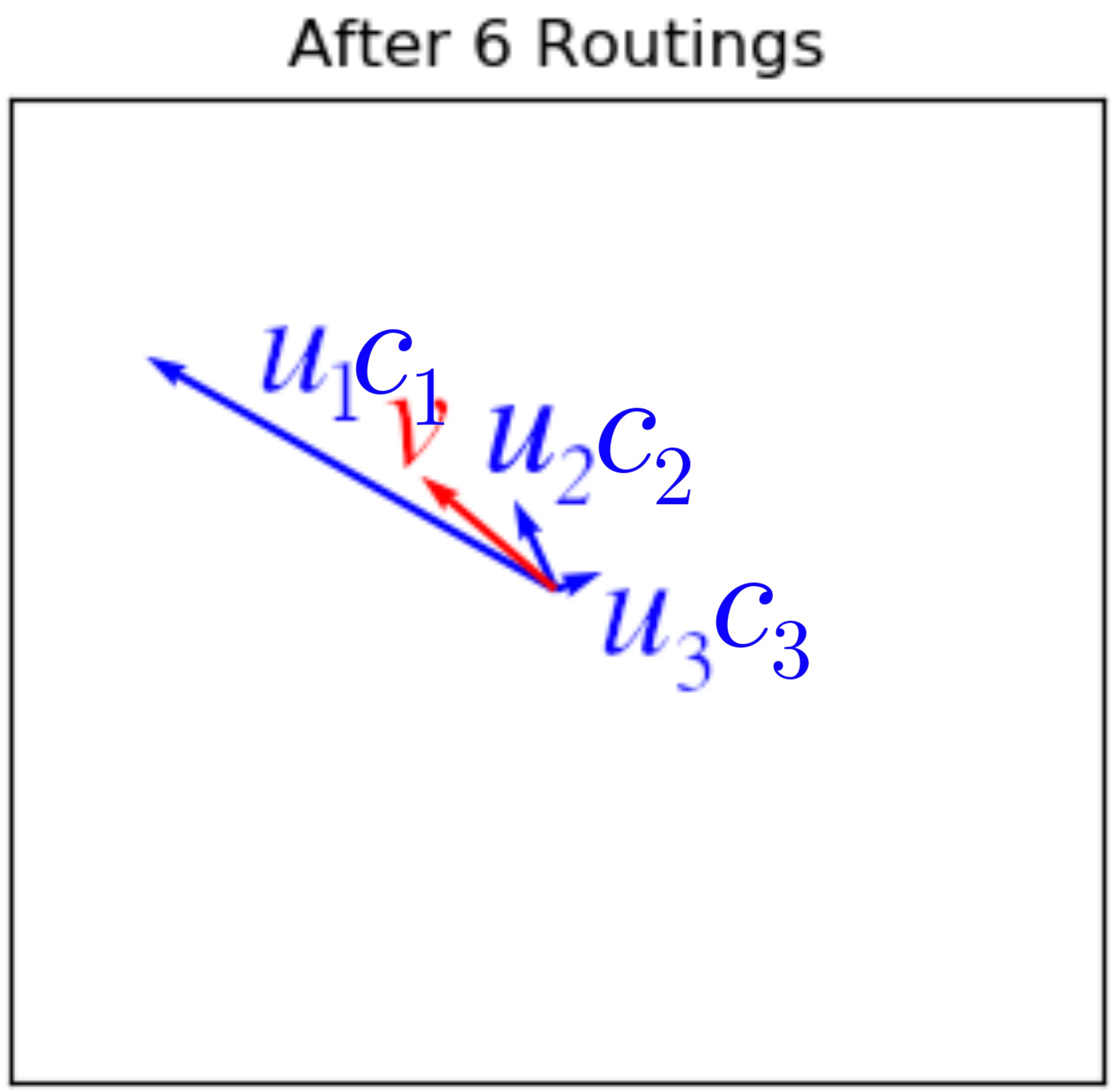
- Calculate weighted average

$$v = \sum_j c_{(j)} u_{(j)}$$

- Update weights $c_{(j)}$ based on $v \cdot u_{(j)}$



Dynamic Routing



Routing by agreement algorithm:

- Calculate weighted average

$$v = \sum_j c_{(j)} u_{(j)}$$

- Update weights $c_{(j)}$ based on $v \cdot u_{(j)}$

- Contributions agreeing with overall consensus preferred
- Longer vectors weighted even more

➔ Enhances agreeing contributions

Activation Function

- Default functions ill suited

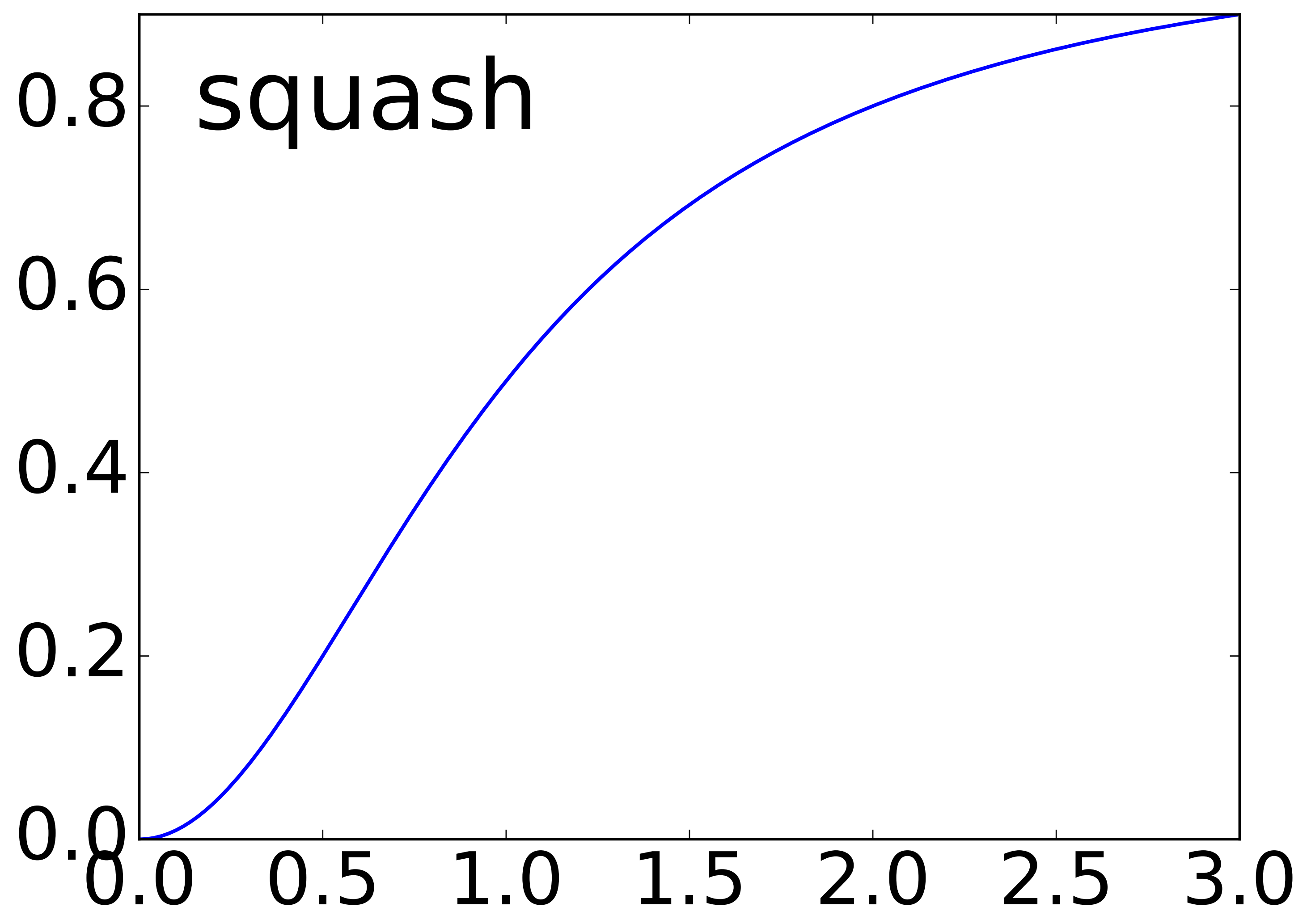
➔ Specialised 'squashing' function:

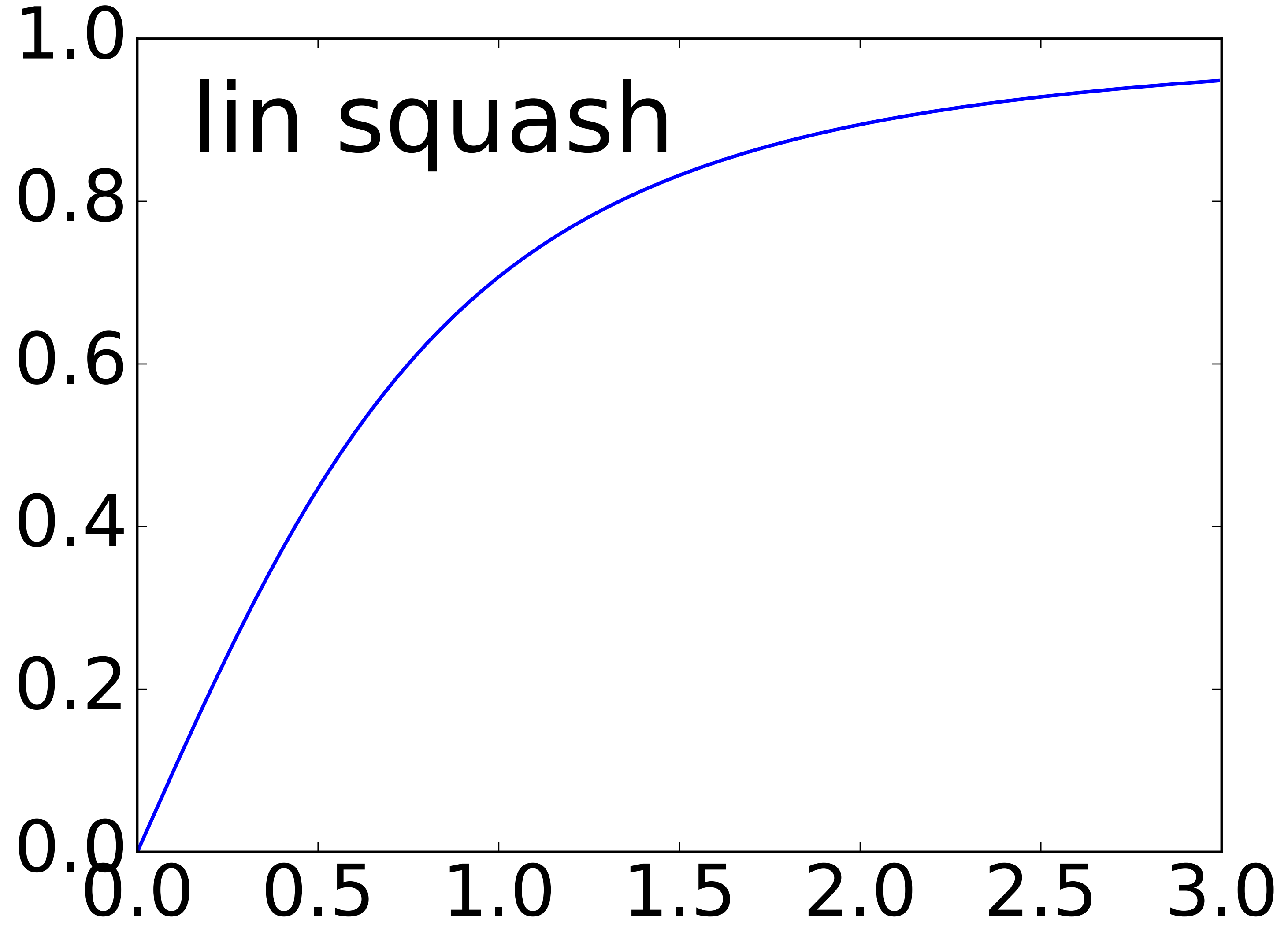
$$\vec{v} \to \vec{v}' = \frac{\vec{v}^2}{1+\vec{v}^2}\hat{v}$$

- Quadratic suppression problematic
  - Value underflow
  - Vanishing gradients

➔ Linearised squashing for deep Caps

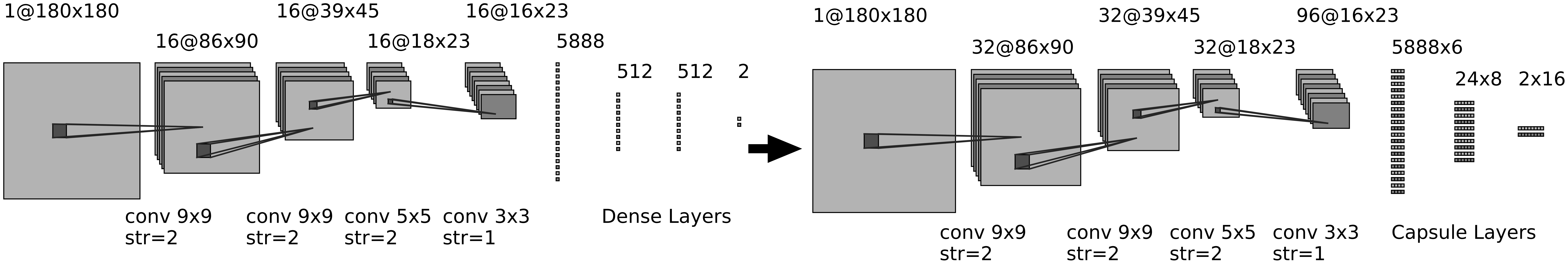
$$\vec{v} \to \vec{v}' = \frac{|\vec{v}|}{\sqrt{1+\vec{v}^2}}\hat{v}$$





CapsNet Architecture

- Convolutions still invaluable for feature extraction
- Replace dense layers of CNN with capsules



‘Conv’ Capsnet

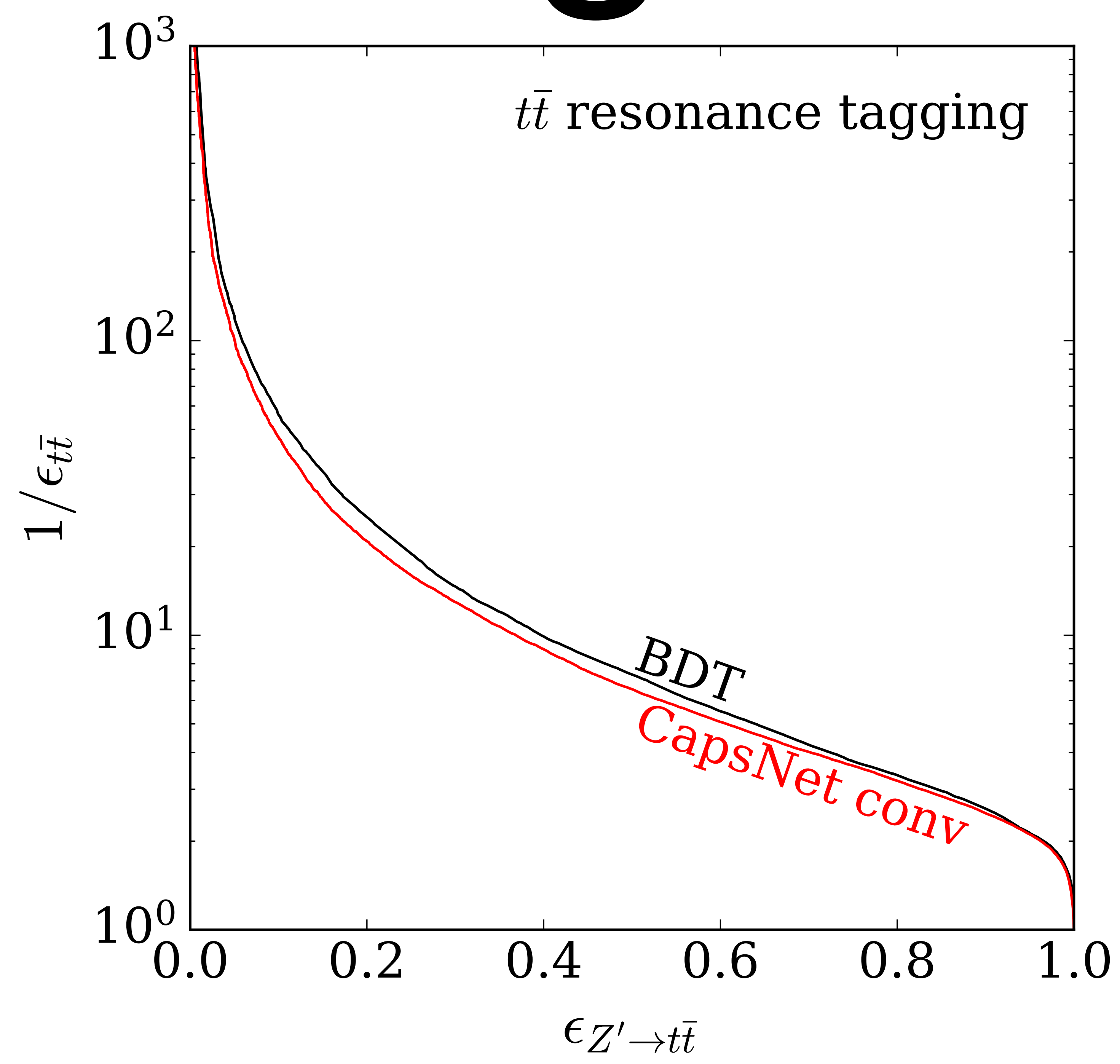
- Number of output capsules equal to classes in dataset
- Length of output capsules are used as classification scores
- Full event input: 180x180 pixels

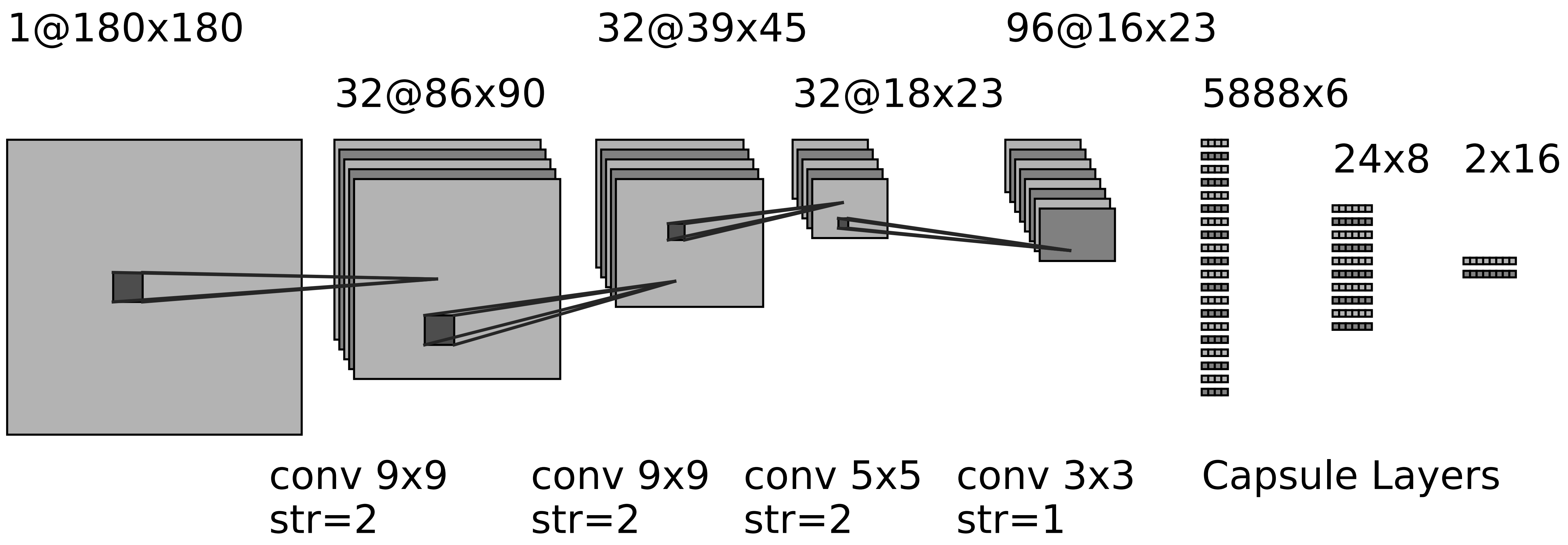
Performance & Insight

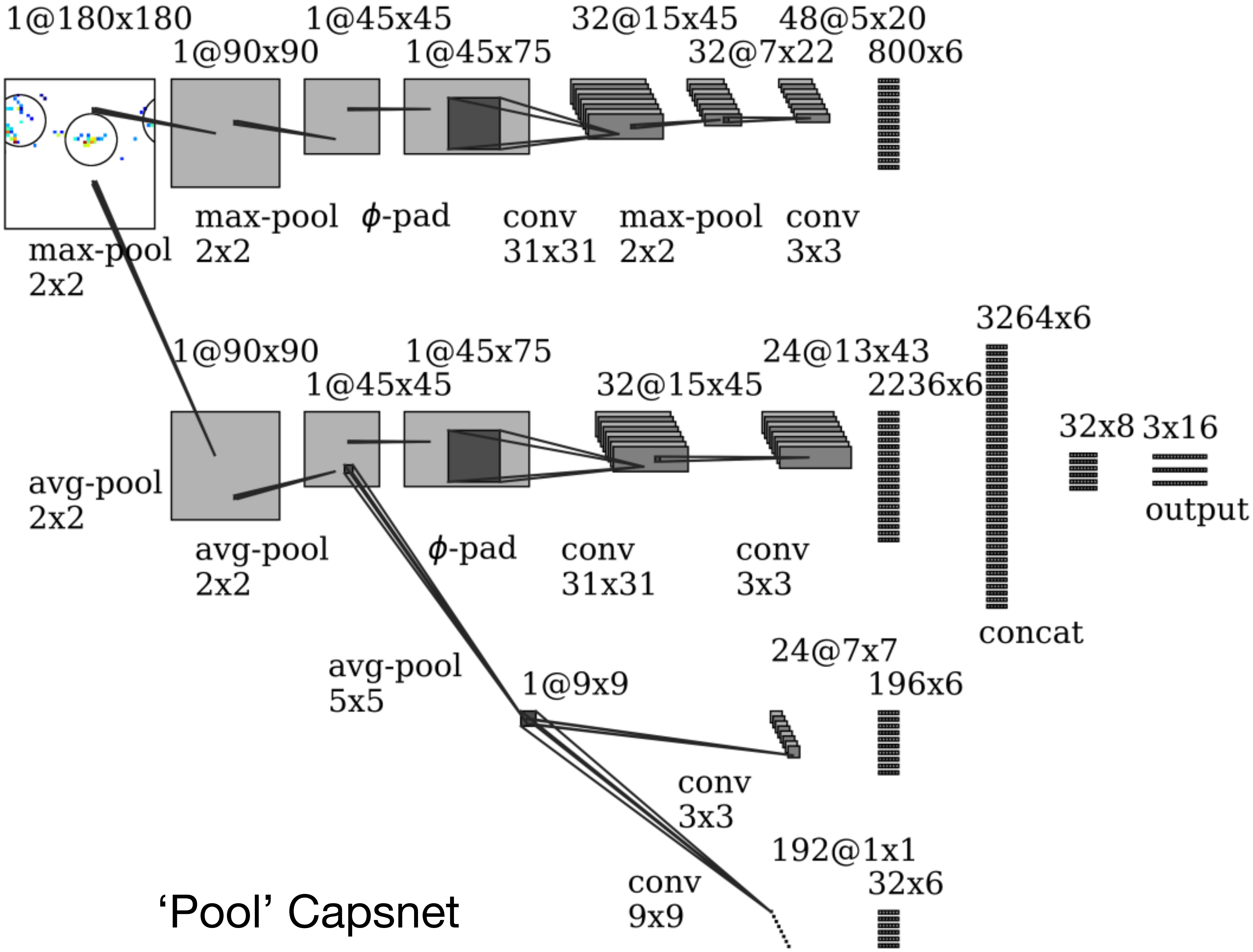
Di-Top Benchmarking

- Initial comparison agains BDT
  - BDT Input: $m_{jj}, p_{T1}, p_{T2}, \eta_1, \eta_2$
- Convolution structure insufficient
  - Can't learn event level features

➔ Specialised architecture needed

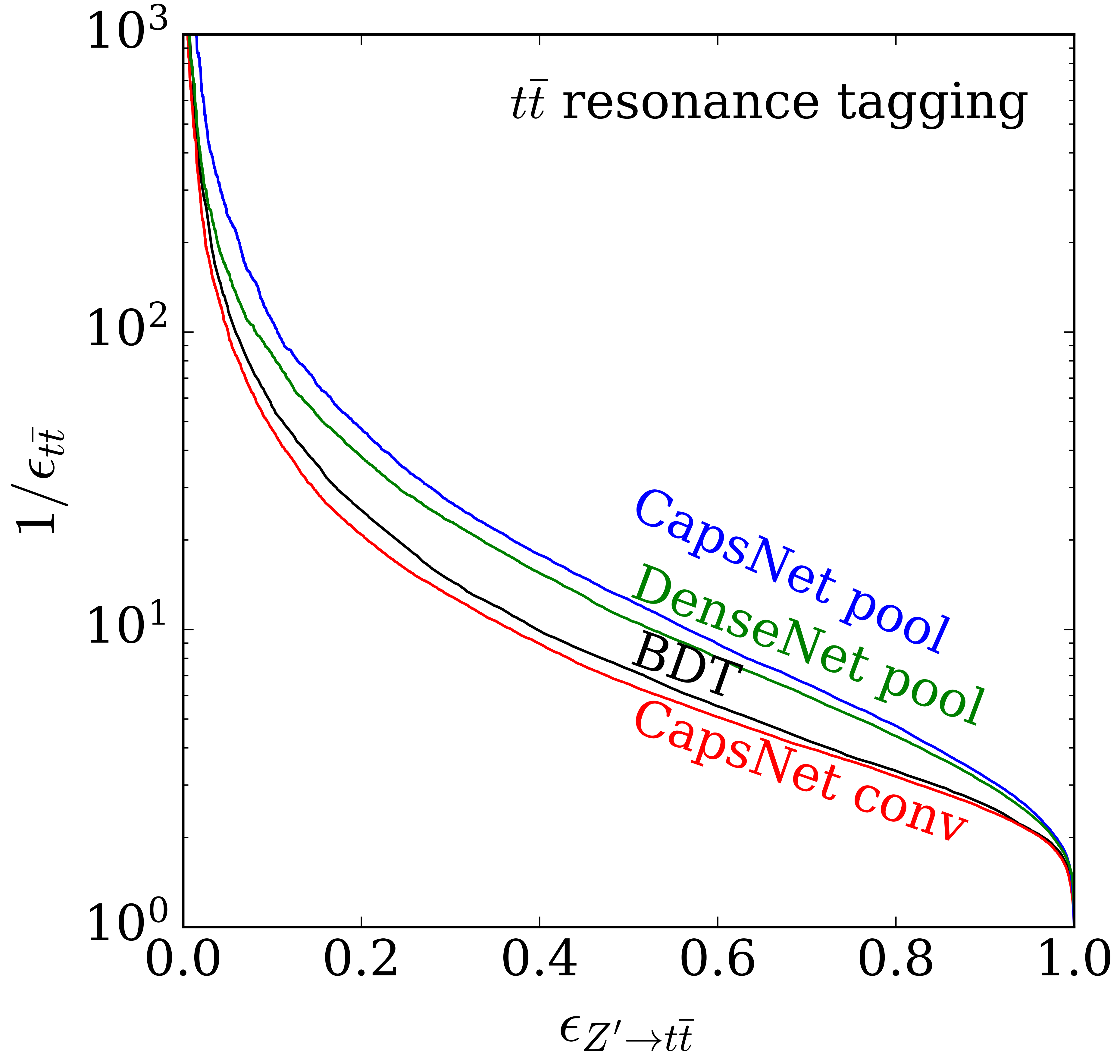






‘Pool’ Capsnet

Di-Top Benchmarking



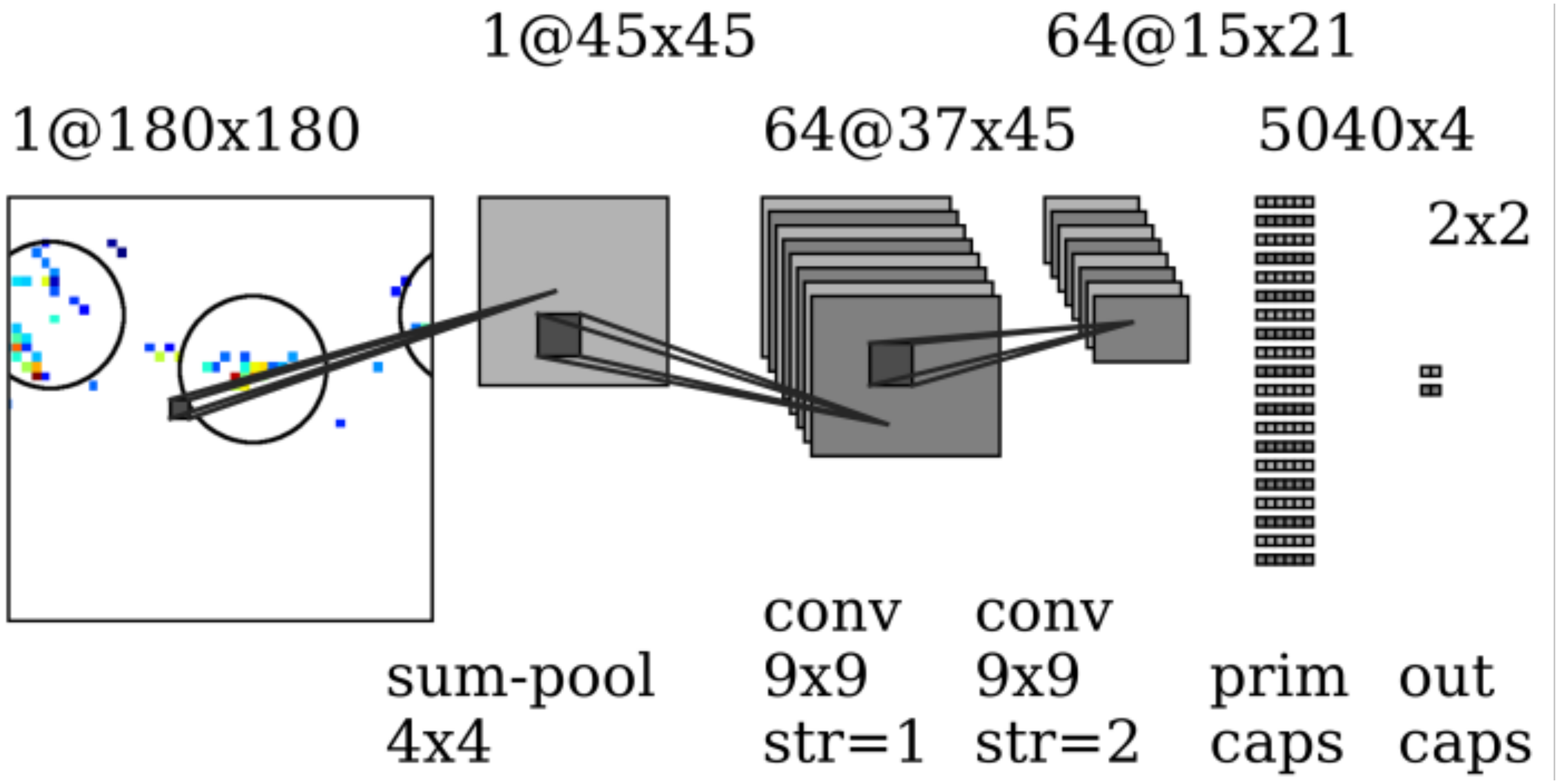
- 'Pool'-Caps architecture beats both BDT and CNN

➡ Capsule network maintain competitive performance

Understanding Capsules

- Proof on concept, performance less important

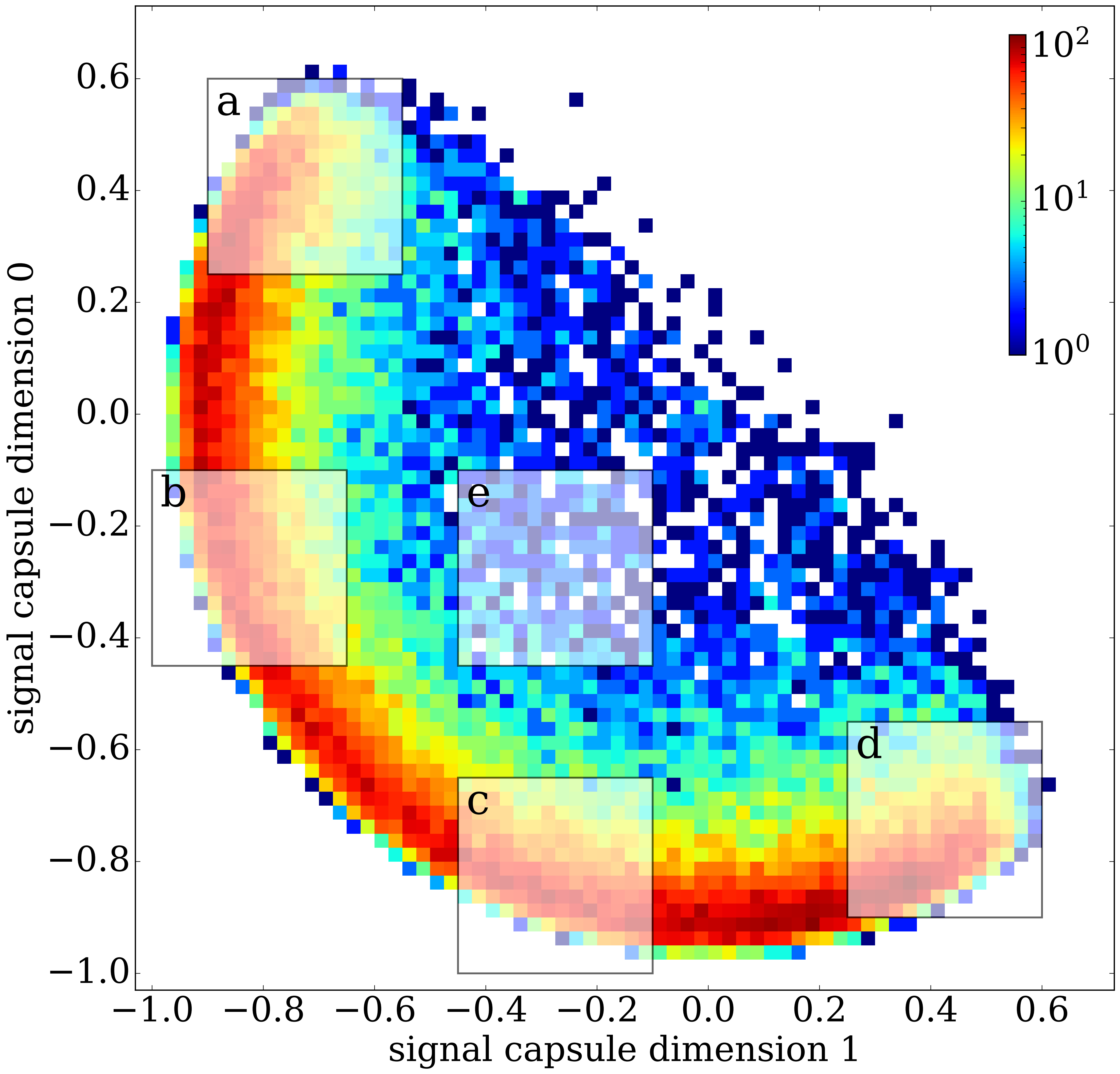
➔ Simplified architecture for ease of understanding



- Most important: Only two-dimensional output capsules

➔ Instantiation vectors plot-able

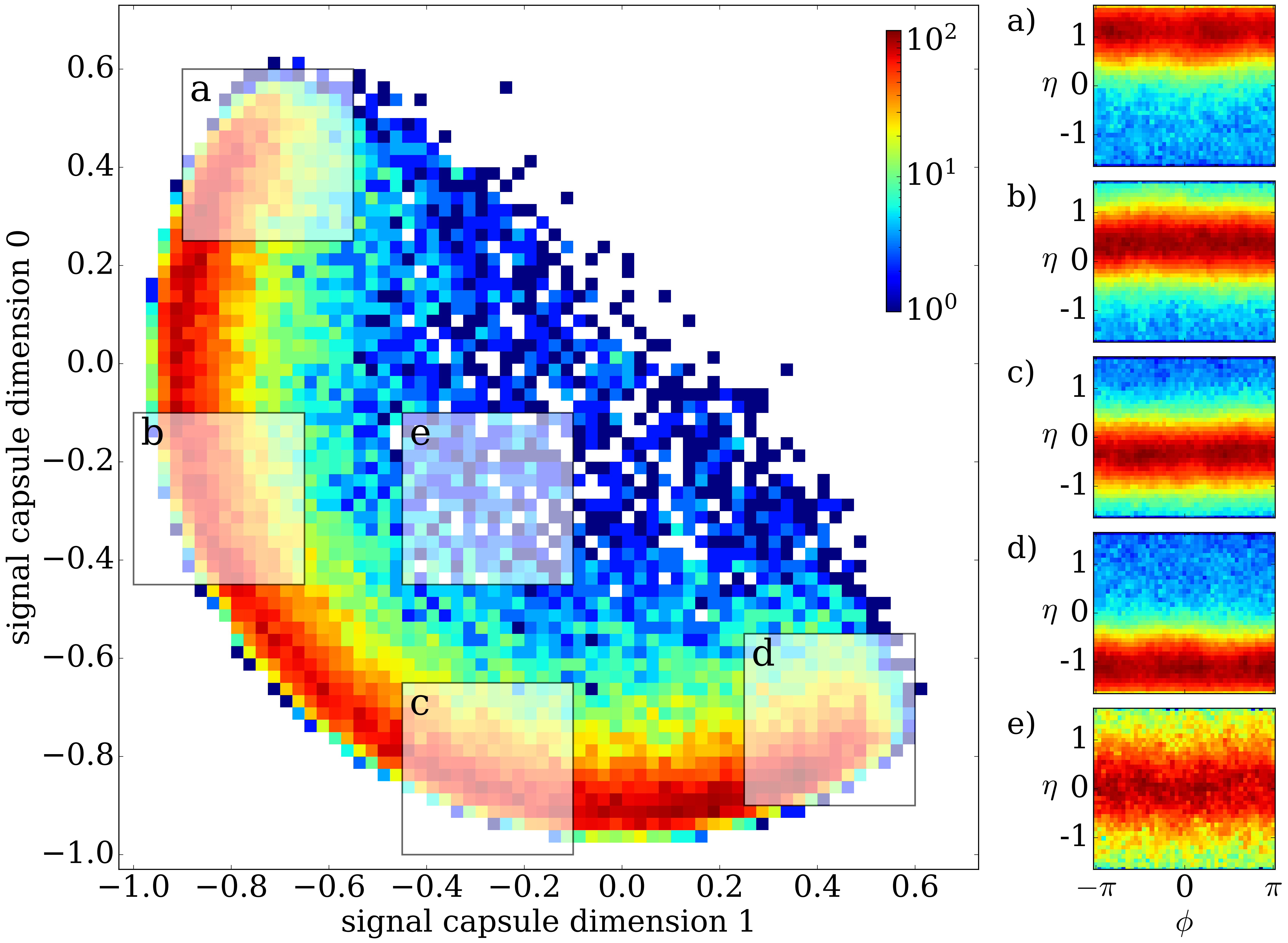
Signal Cap for Signal



- Capsule outputs plotted in 2d
- Vector length ~ 1.0

➔ Correct prediction

Signal Cap for Signal

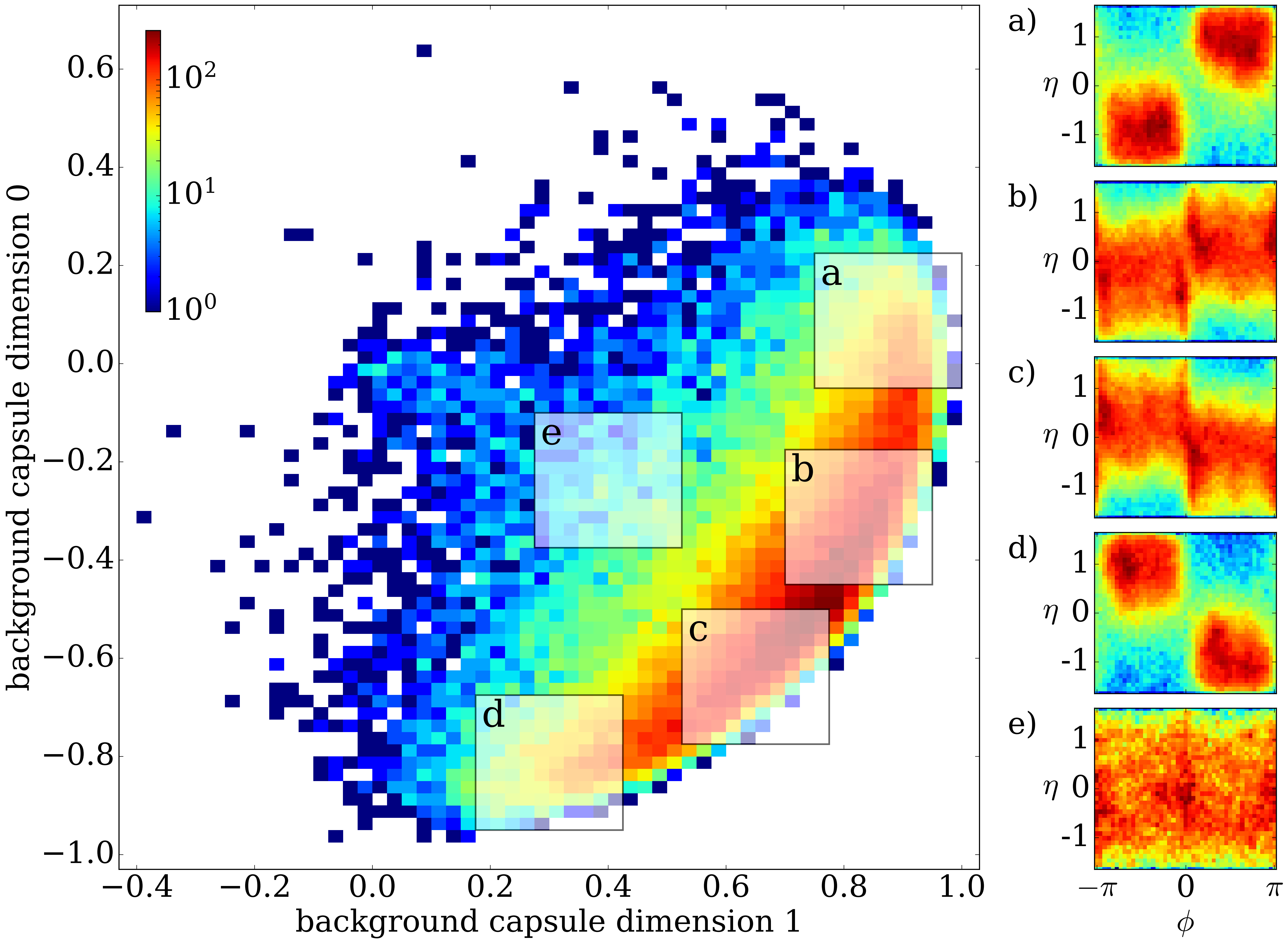


- Capsule outputs plotted in 2d
- Vector length ~ 1.0

  ➔ Correct prediction
- Map back to event images
- Tops form Z' decay are back-to-back in Z' rest frame

  ➔ Both jets have similar η

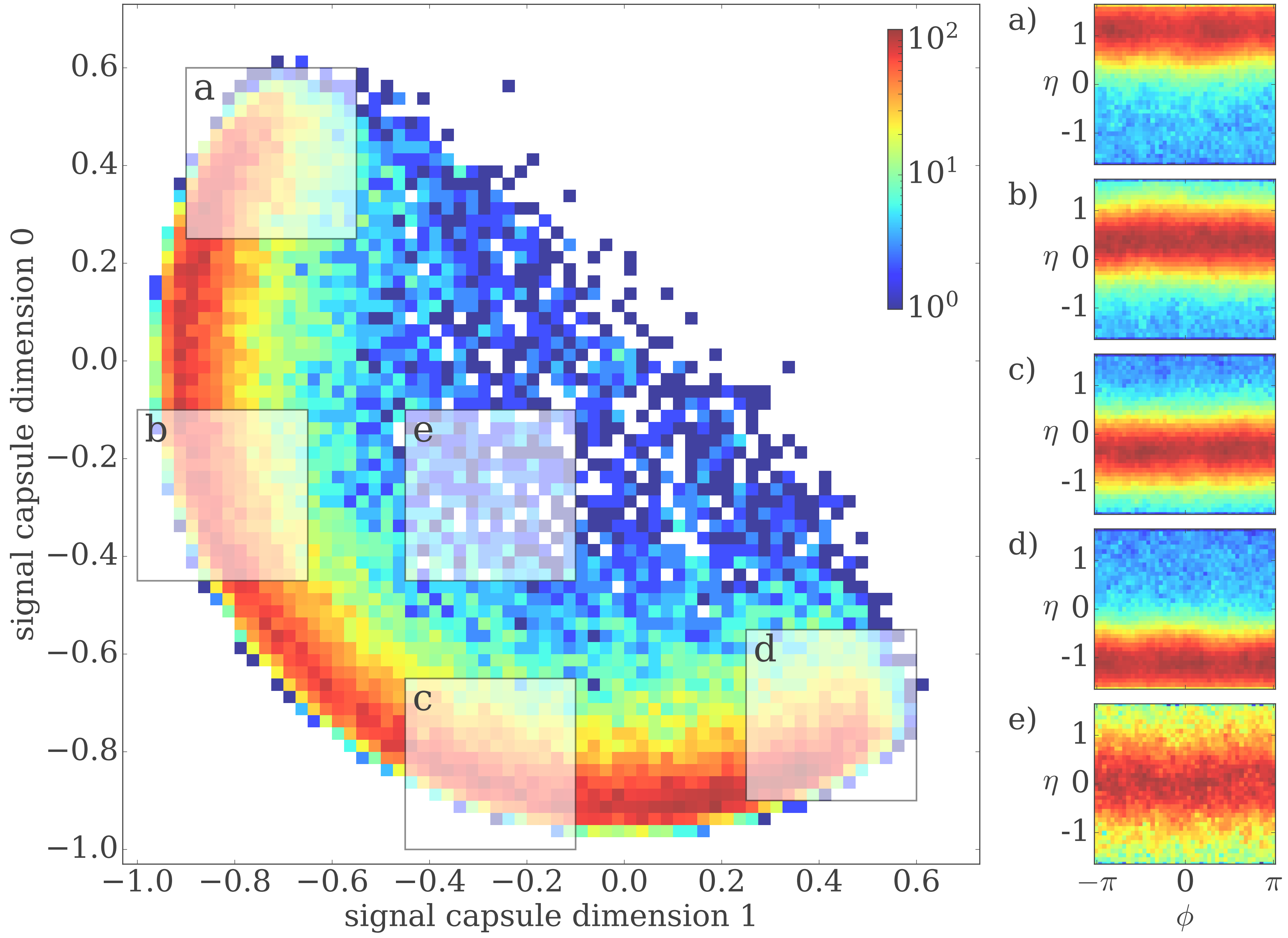
Background Cap for Background



- Capsule outputs plotted in 2d
- Vector length ~ 1.0
  - ➔ Correct prediction
- Map back to event images
- Tops form Z' decay are back-to-back in Z' rest frame
  - ➔ Both jets have similar η
- QCD jets are back to back in detector frame
  - ➔ Jets have 'opposite' position
  - ➔ Angle correlated to physics

Conclusion

- Capsules show great potential for explainability
  - Maintains CNN performance
  - Can handle whole event inputs
- Multi-classifier for Z', di-top, di-light-jet
  - Works better than Z' vs (tt + light light)
- High activity events
  - Works on $t\bar{t}H$
- Full paper 1906.11265



Thank you