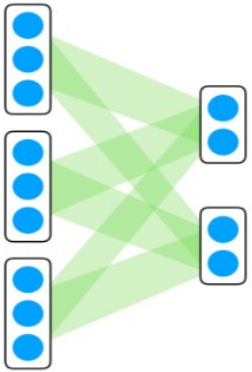
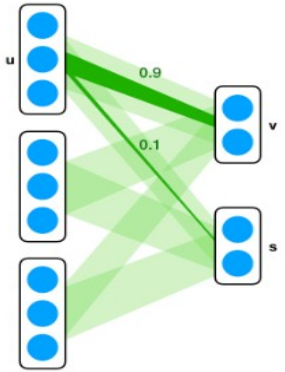
Analyzing the Explanation and Interpretation Potential of Matrix Capsules





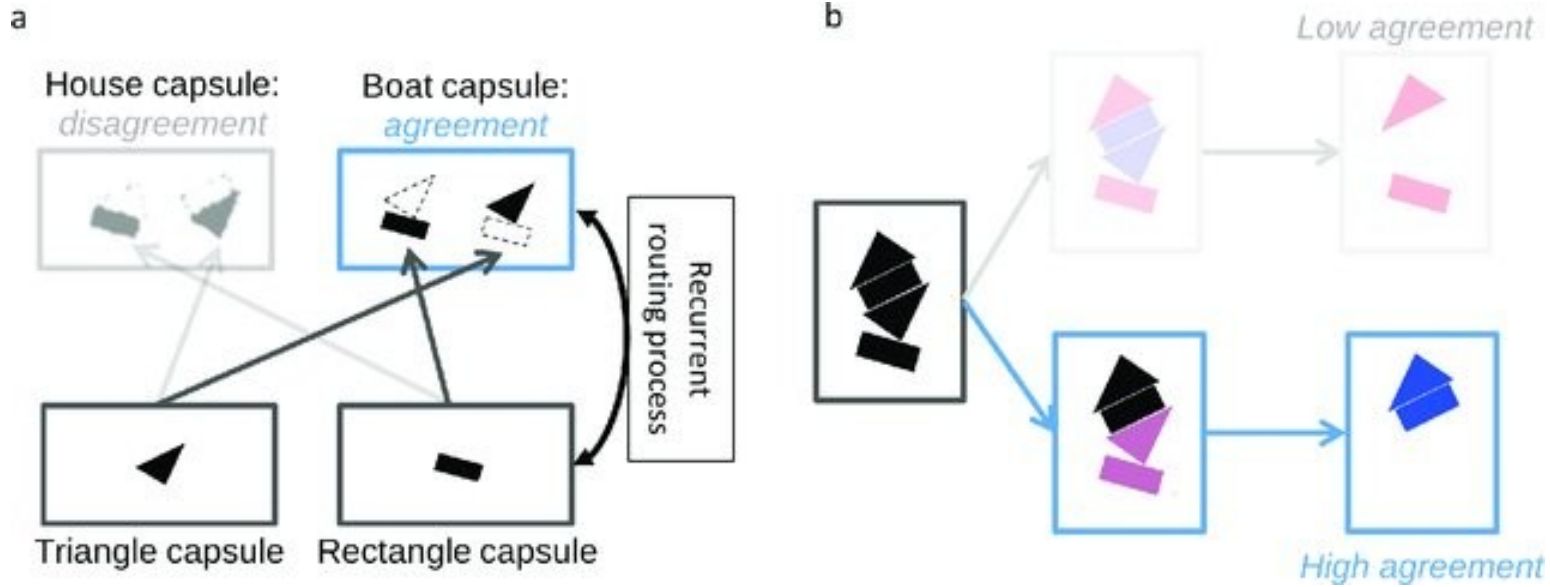
Andrei Bondarenko, **Saja AL-Tawalbeh**, José Oramas
University of Antwerp - imec-IDLab

Outline

- Introduction
- Motivation
- Methodology
- Results
- Conclusions

Introduction

- Capsule Networks

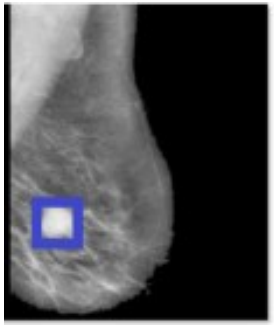


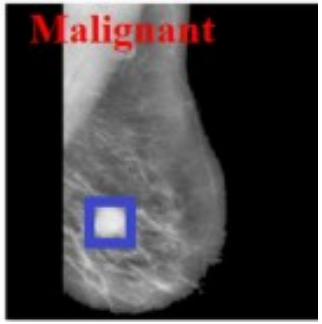
Illustrations by: Novatec Consulting GmbH

Motivation

Diabetic detection

Breast Cancer Diagnosis





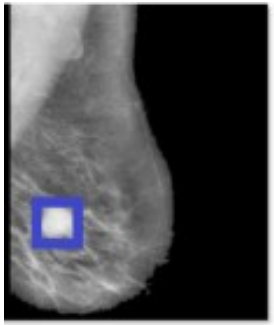
Automatic Target Recognition

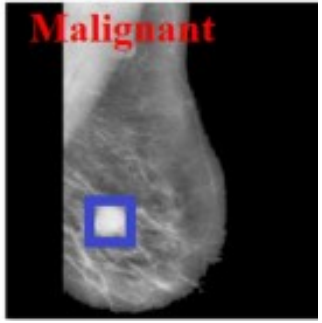
1- Kalyani, et al. "Diabetic retinopathy detection and classification using capsule networks." Complex & Intelligent Systems (2021)
2- Anupama, et al. "Breast cancer classification using capsule network with preprocessed histology images." International conference on communication and signal processing (2019)
3- Shah, et al. "Automatic target recognition from SAR images using capsule networks." Pattern Recognition and Machine Intelligence (2019)

Motivation

Diabetic detection

Breast Cancer Diagnosis





Automatic Target Recognition

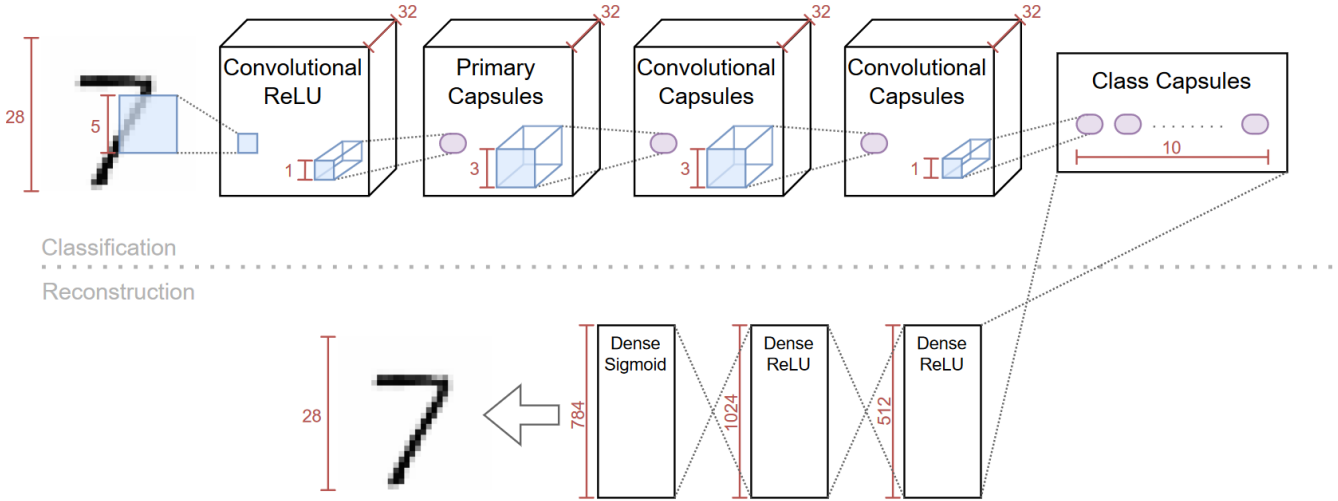
Existing works focus on original capsule architecture, newer architectures barely studied!

1- Kalyani, et al. "Diabetic retinopathy detection and classification using capsule networks." Complex & Intelligent Systems (2021)
2- Anupama, et al. "Breast cancer classification using capsule network with preprocessed histology images." International conference on communication and signal processing (2019)
3- Shah, et al. "Automatic target recognition from SAR images using capsule networks." Pattern Recognition and Machine Intelligence (2019)

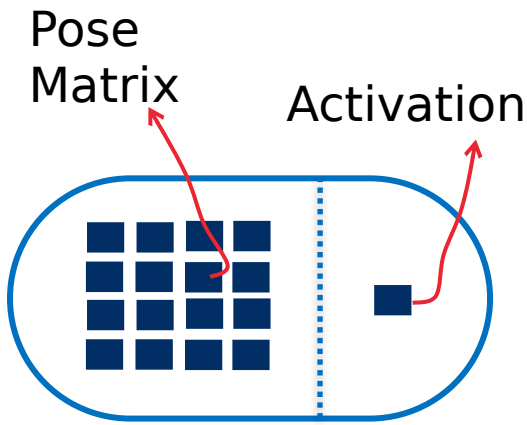
How?

How did we attempt solve the problem?

Study of Matrix Capsules with EM-Routing



Architecture

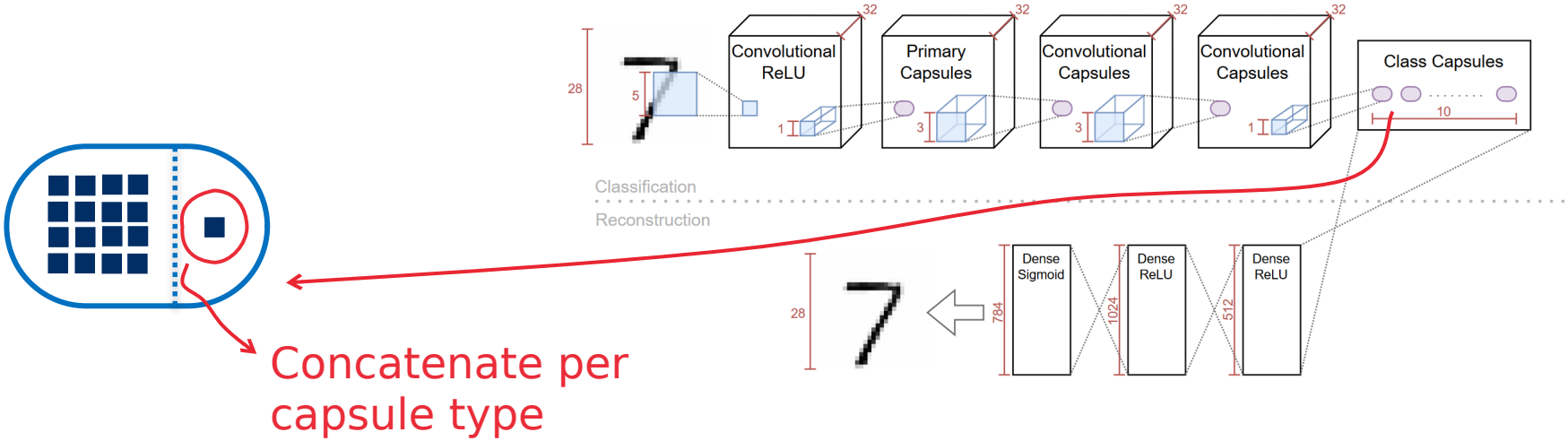


How do they differ?

Hinton, Geoffrey E., Sara Sabour, and Nicholas Frosst. "Matrix capsules with EM routing." ICLR. 2018.

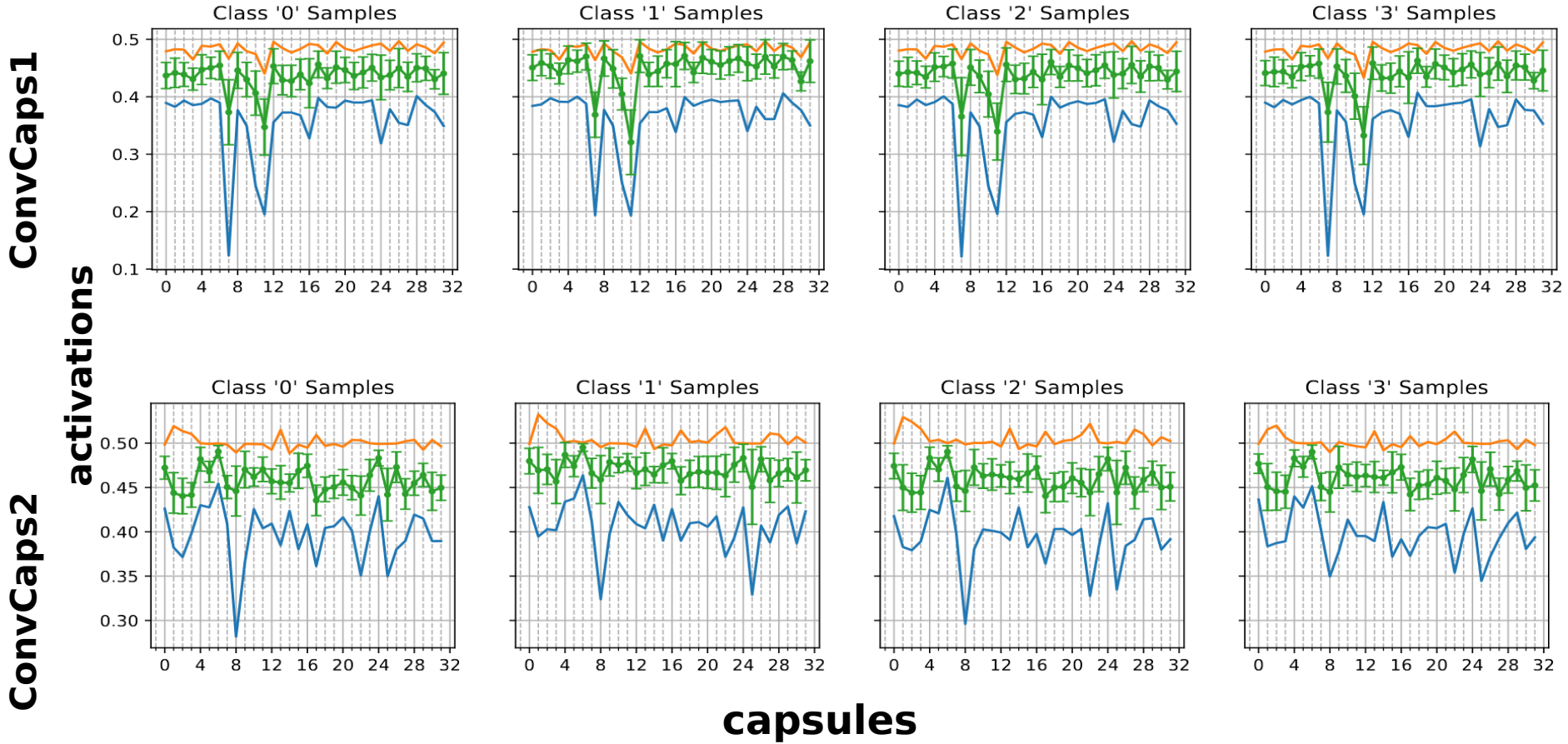
Methodology

Activations

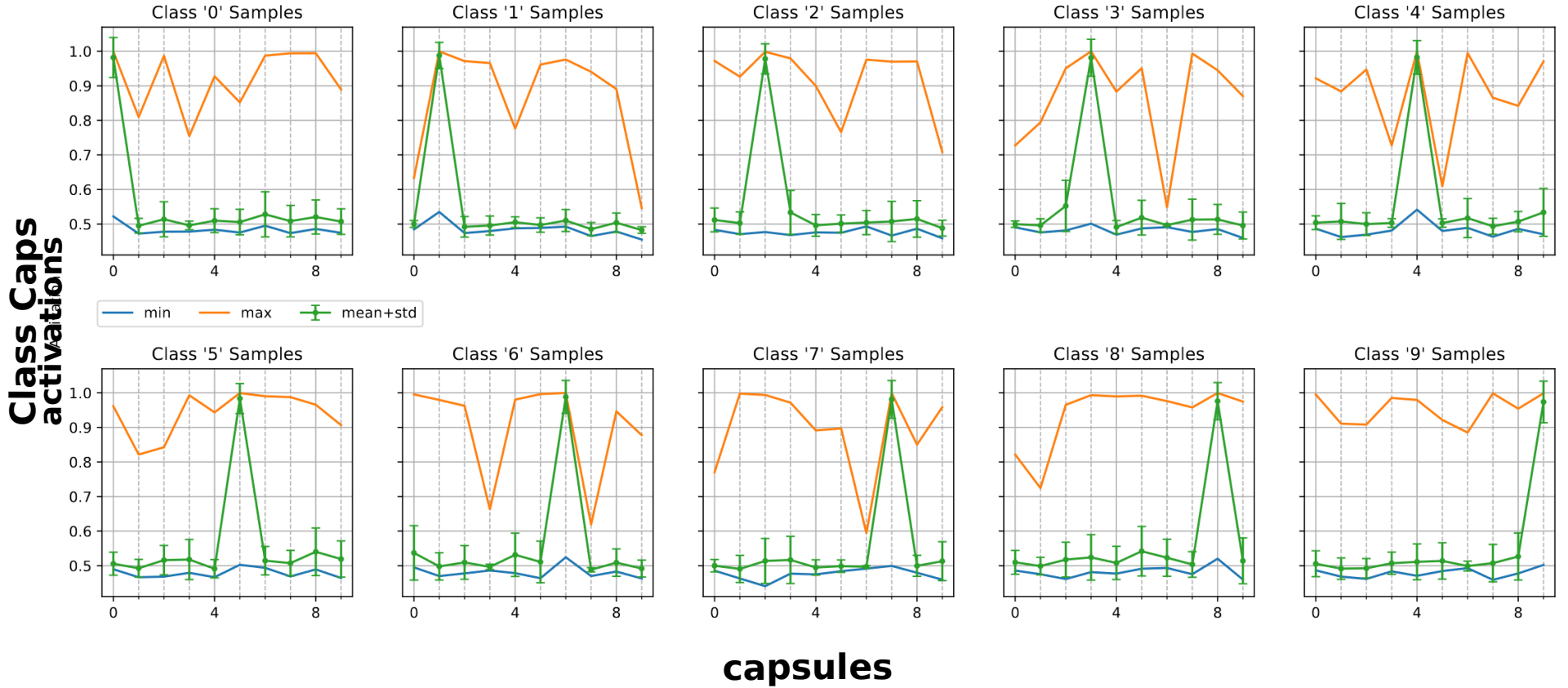


Concatenate per capsule type

Activations

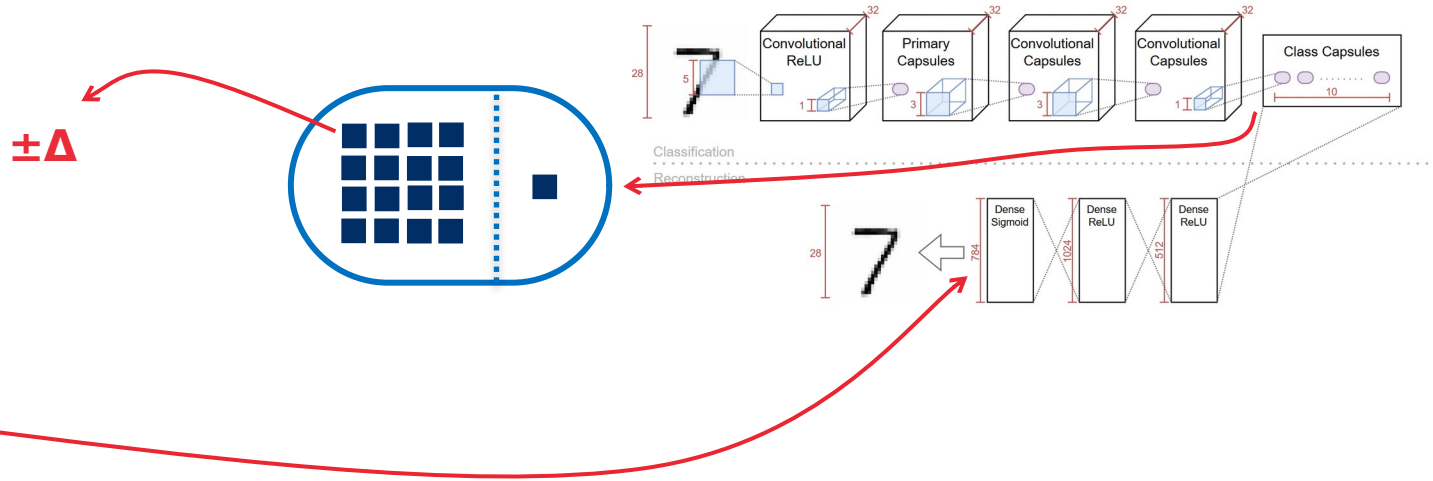


Activations

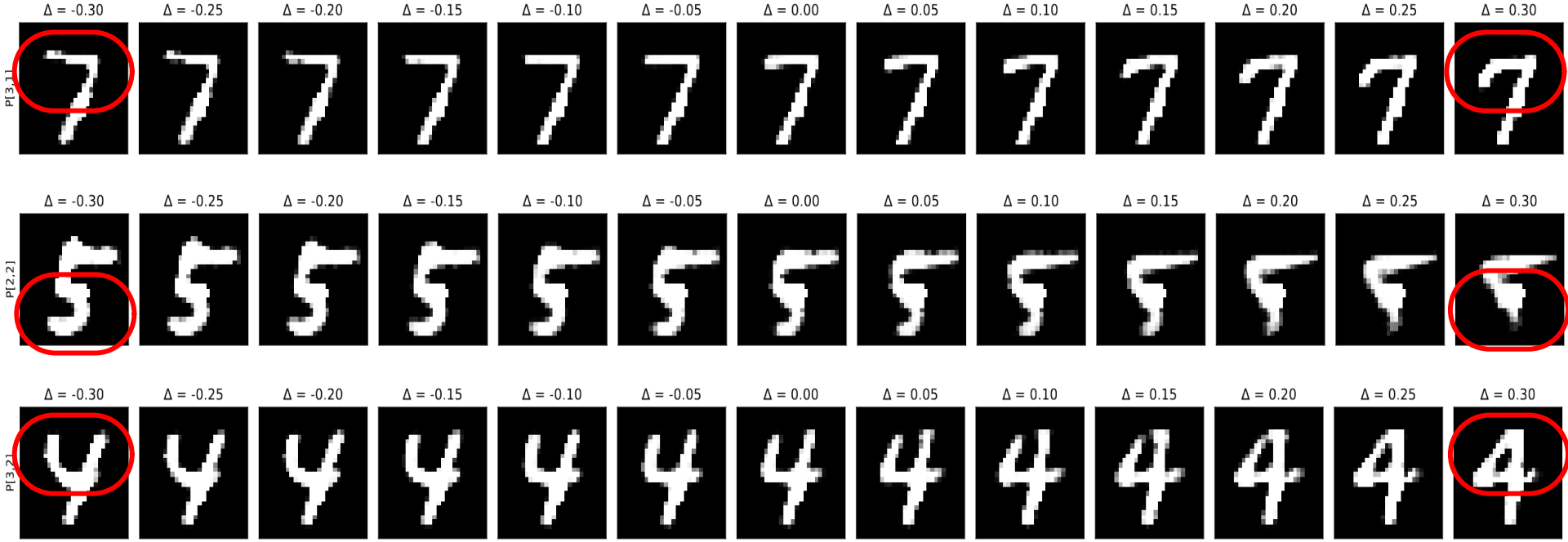


Pose Matrix Perturbations

- Interval:
  - [-0.3,+0.3]
- Step:
  - [0.05]

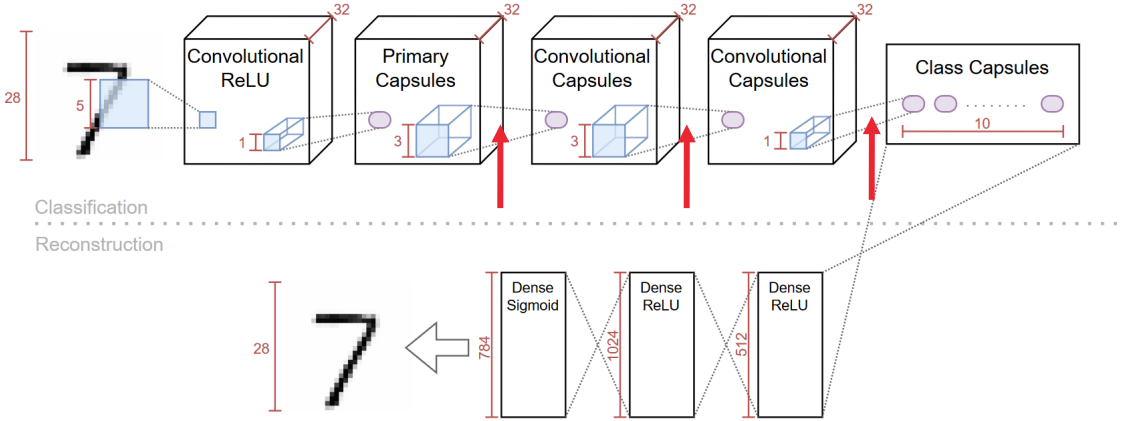


Results / Pose Matrix Perturbations



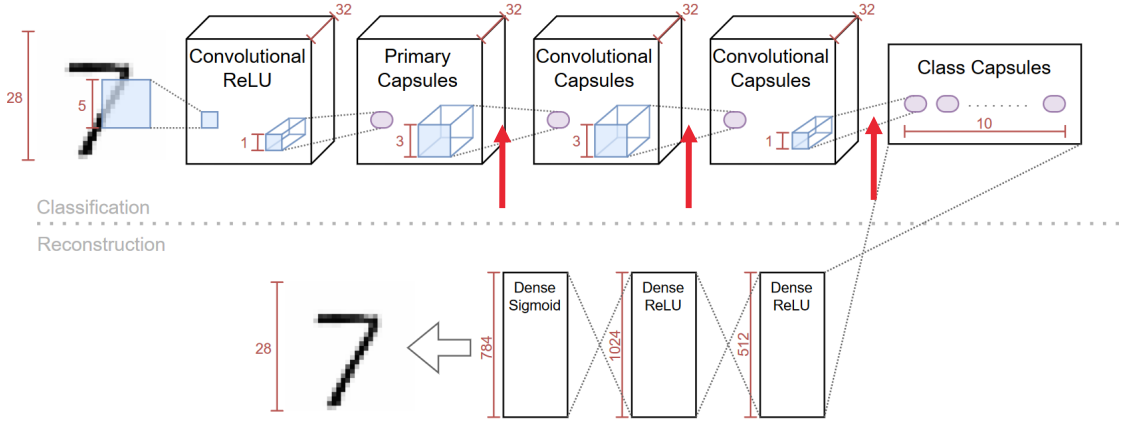
Capsule Importance via Routing Coefficients

- Per capsule layer
- Retrieve routing coefficients



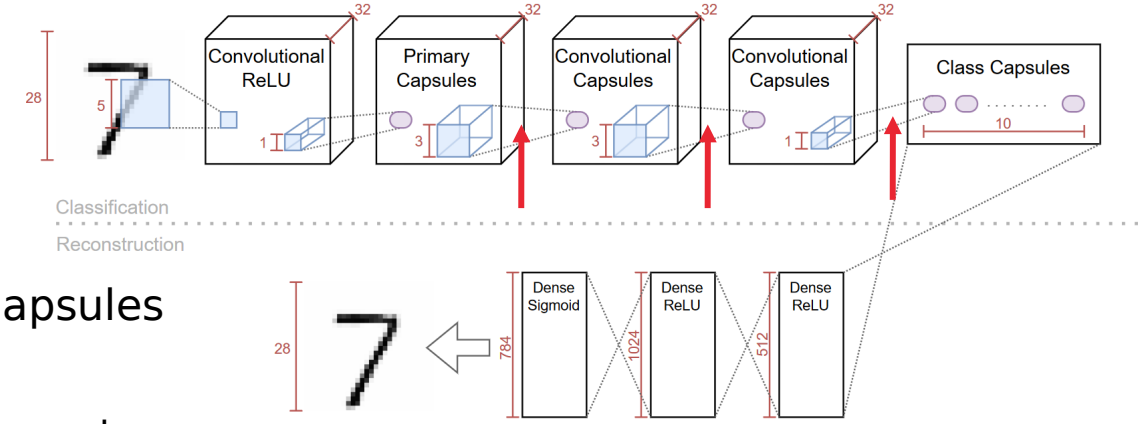
Capsule Importance via Routing Coefficients

- Per capsule layer
- Retrieve routing coefficients
- Measure of importance

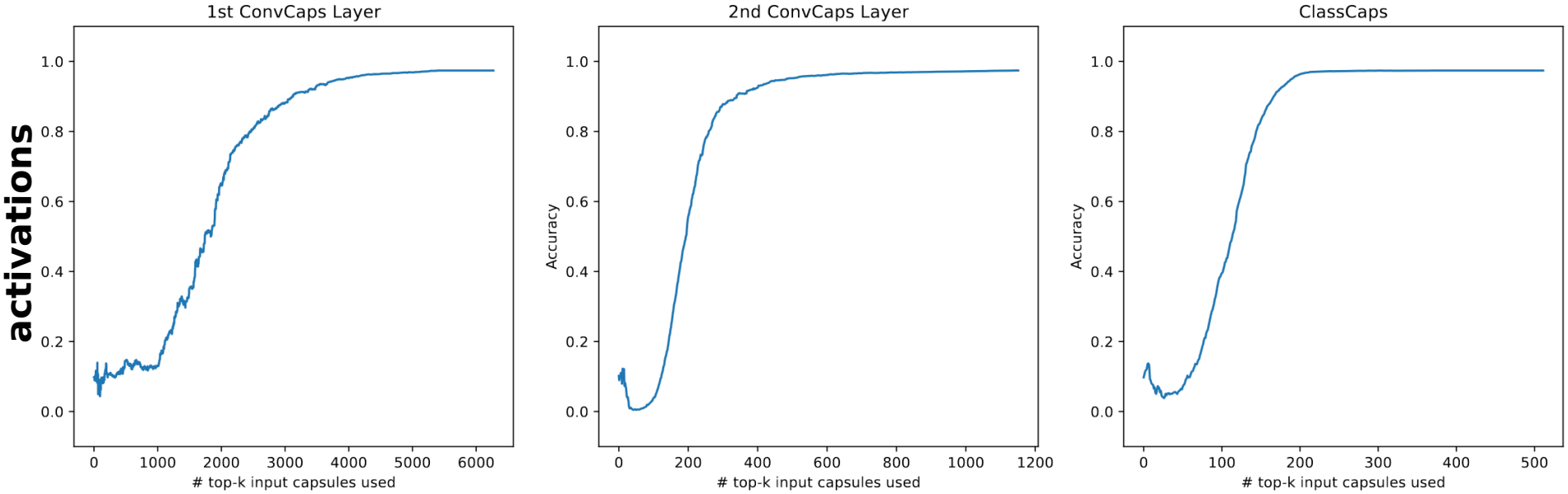


Capsule Importance via Routing Coefficients

- Per capsule layer
- Retrieve routing coefficients
- Measure of importance
- Compute ranking of relevant capsules per layer:
  - Masking the less relevant capsules

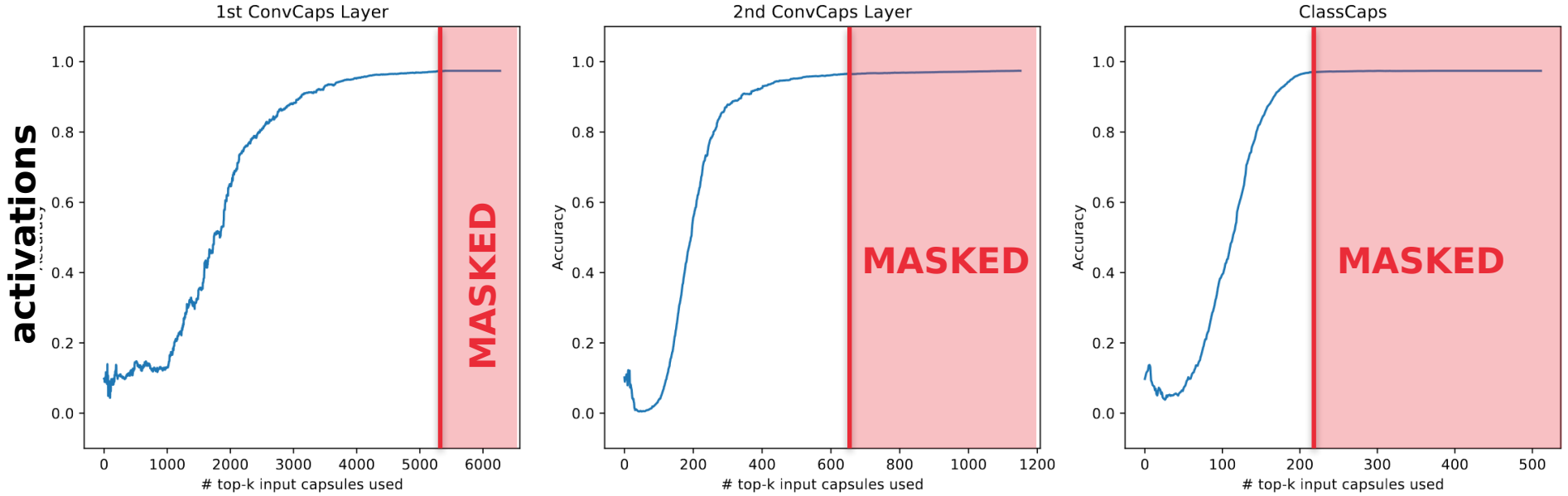


Capsule Importance via Routing Coefficients



Top-k relevant Capsules

Capsule Importance via Routing Coefficients



Top-k relevant Capsules

Capsule Importance via Routing Coefficients

Masked			Classification Performance (%)
ConvCaps1	ConvCaps2	ClassCaps	
≈5000	≈600	≈200	-
			97
X			97
	X		96
		X	97
X	X		96
X		X	96
	X	X	92
X	X	X	91

Conclusions and Future work

- The outcomes are preliminary
- Pose matrices
  - Verified that parameters encode characteristics

Conclusions and Future work

- The outcomes are preliminary
- Pose matrices
  - Verified that parameters encode characteristics
- Activations
  - Small difference between classes
  - Probably not enough to explain classification
- Routing coefficients
  - Can be used as a proxy for input capsule importance
  - Eliminating less important capsules

Thank You
Questions?