Variational Capsule Encoder

Harish RaviPrakash, Syed Muhammad Anwar, Ulas Bagci

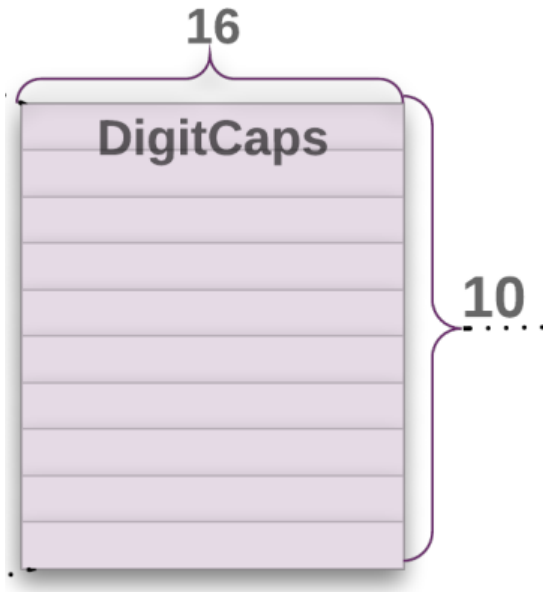
Department of Computer Science, University of Central Florida, Orlando, Florida
Email: harishr@knights.ucf.edu

MOTIVATION

- Variational autoencoders
  - help learn the variations in the different classes in the data.
  - but are not viewpoint invariant.
- Capsule networks
  - Learn a relationship between objects and its parts, thus learning a transformation invariant model.
- *Capsule networks in representation learning has not been examined under the VAE settings*

CAPSULES FOR REPRESENTATION LEARNING

- Capsules replace scalar neurons in networks with vectors
  - Helps implicitly capture information such as orientation
- A capsule represents a feature, and the vector represents its likelihood
- Capsules helps build a hierarchical relationship → part-whole relationship



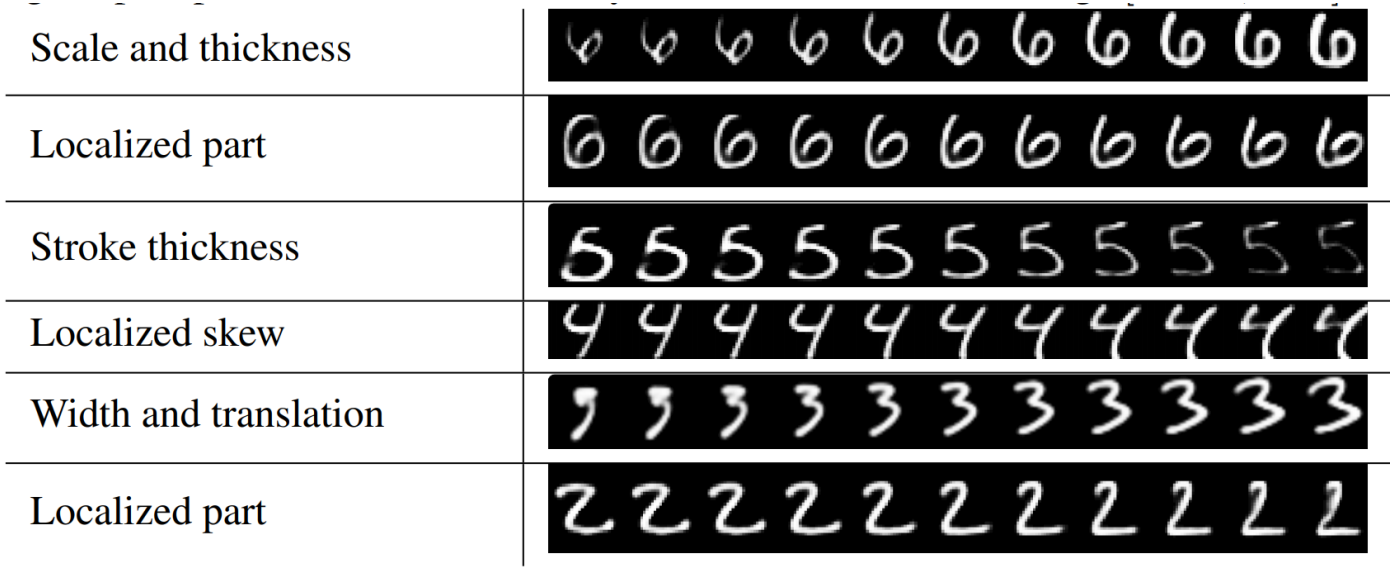
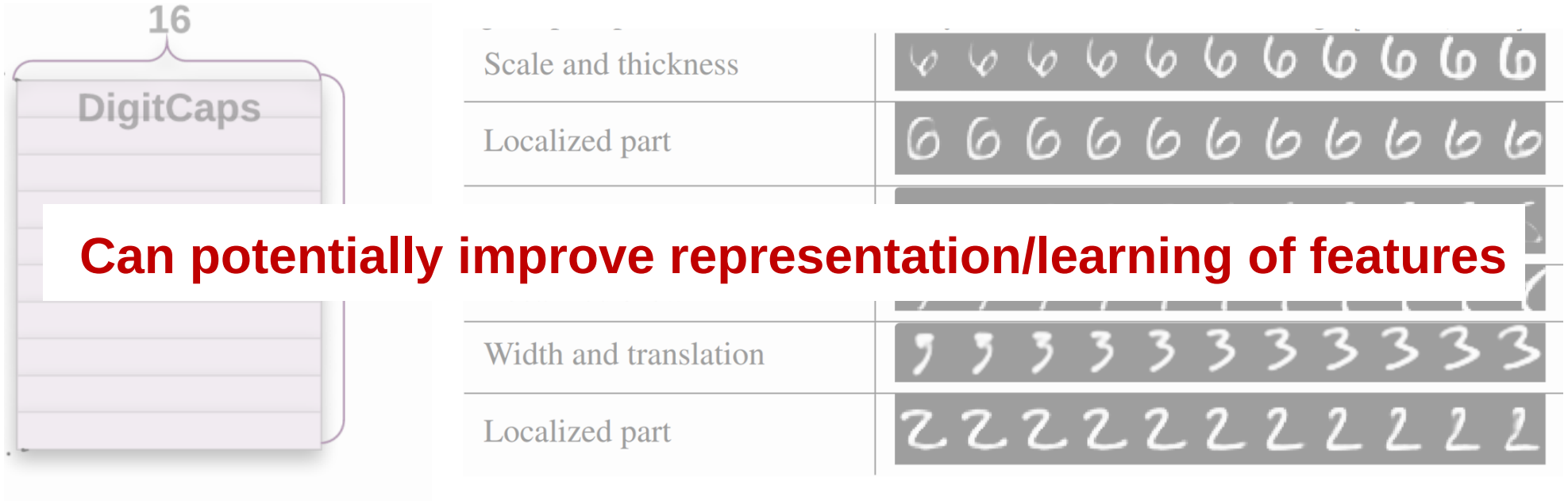


Fig credit: Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic routing between capsules." Advances in neural information processing systems. 2017.

CAPSULES FOR REPRESENTATION LEARNING

- Capsules replace scalar neurons in networks with vectors
  - Helps implicitly capture information such as orientation
- A capsule represents a feature, and the vector represents its likelihood

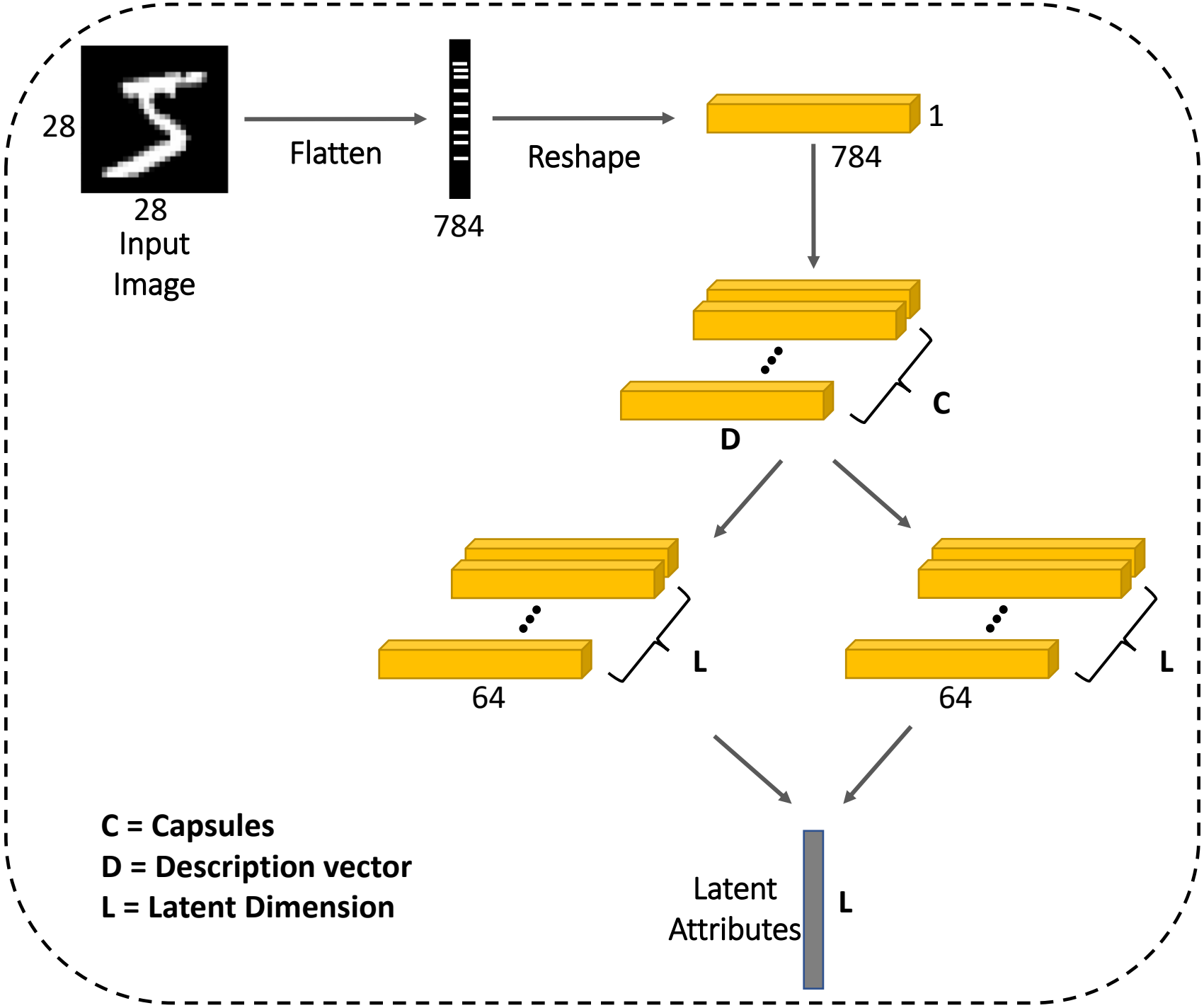


Can potentially improve representation/learning of features

Fig credit: Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic routing between capsules." Advances in neural information processing systems. 2017.

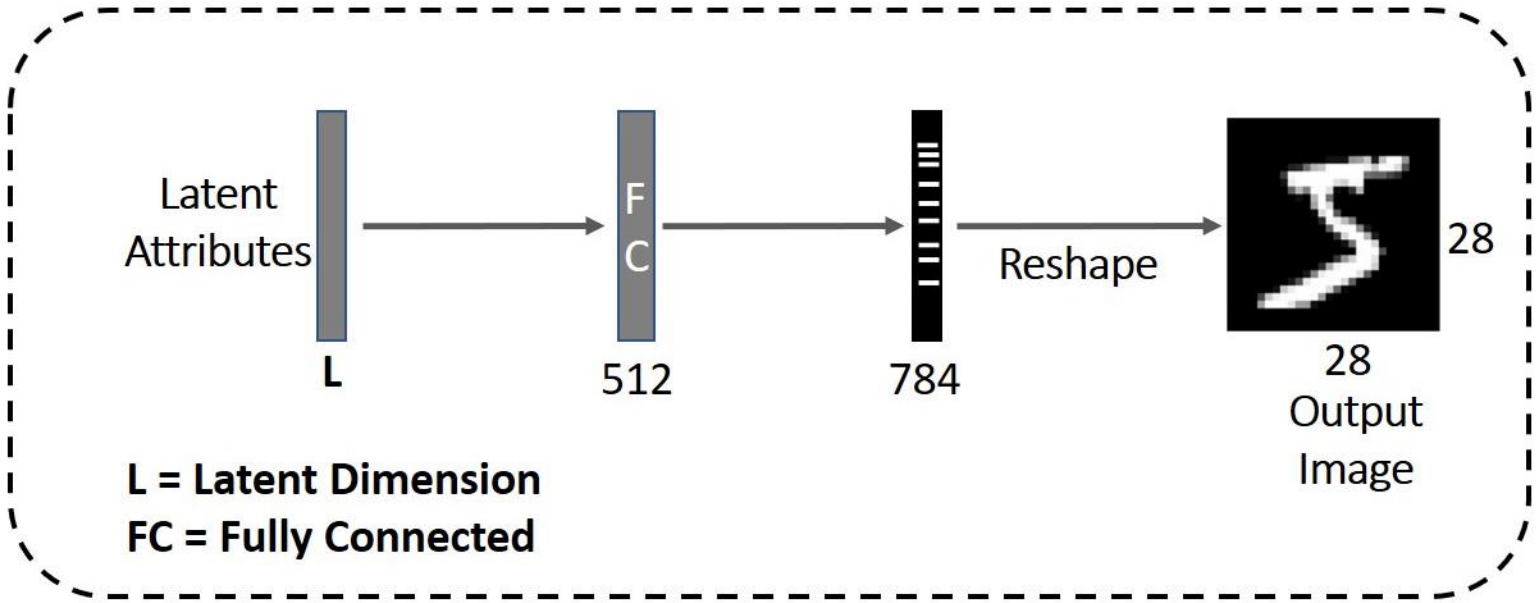
PROPOSED APPROACH

- Novel capsule based variational capsule encoder (B-Caps)
  - Fully connected layers in the encoder are replaced with capsule layer
  - Capsule layers, mean and variance, influence the latent space sampling


28
28
Input
Image
Flatten
784
Reshape
1
784
C
D
L
64
L
64
L
Latent
Attributes
C = Capsules
D = Description vector
L = Latent Dimension

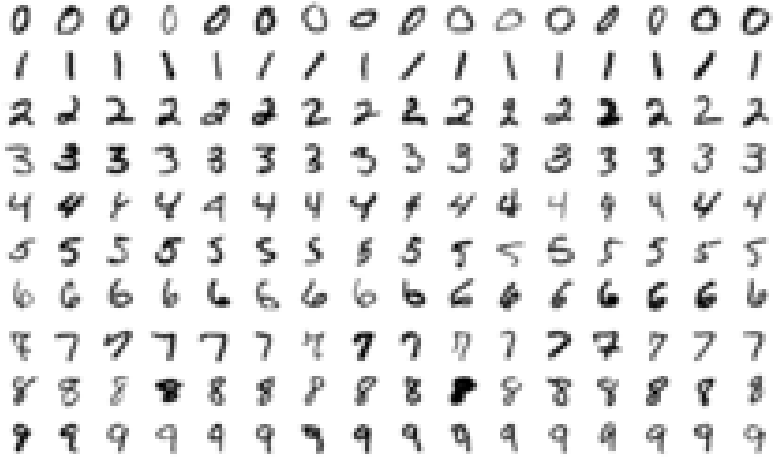
PROPOSED APPROACH

- Decoder network is from the basic VAE with fully connected layers.



EXPERIMENTS - DATA

- MNIST dataset
  - 60000 training images – 10 classes/digits
  - 10000 test images



- FASHION-MNIST dataset
  - 60000 training images – 10 classes
  - 60000 test images

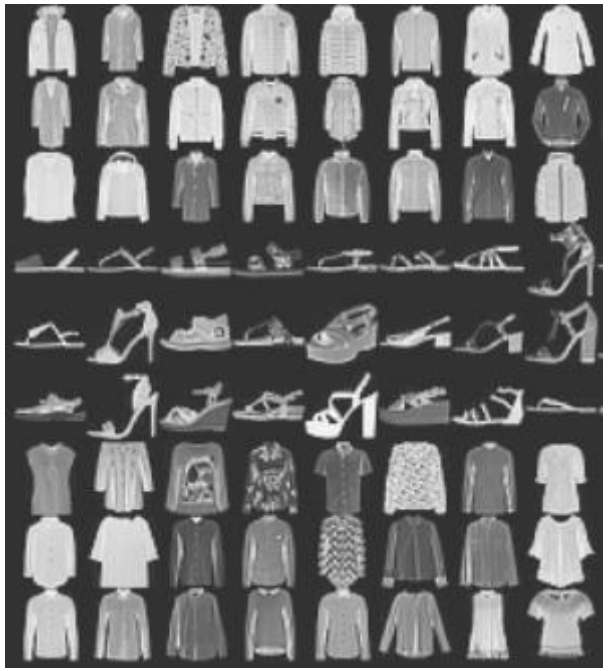
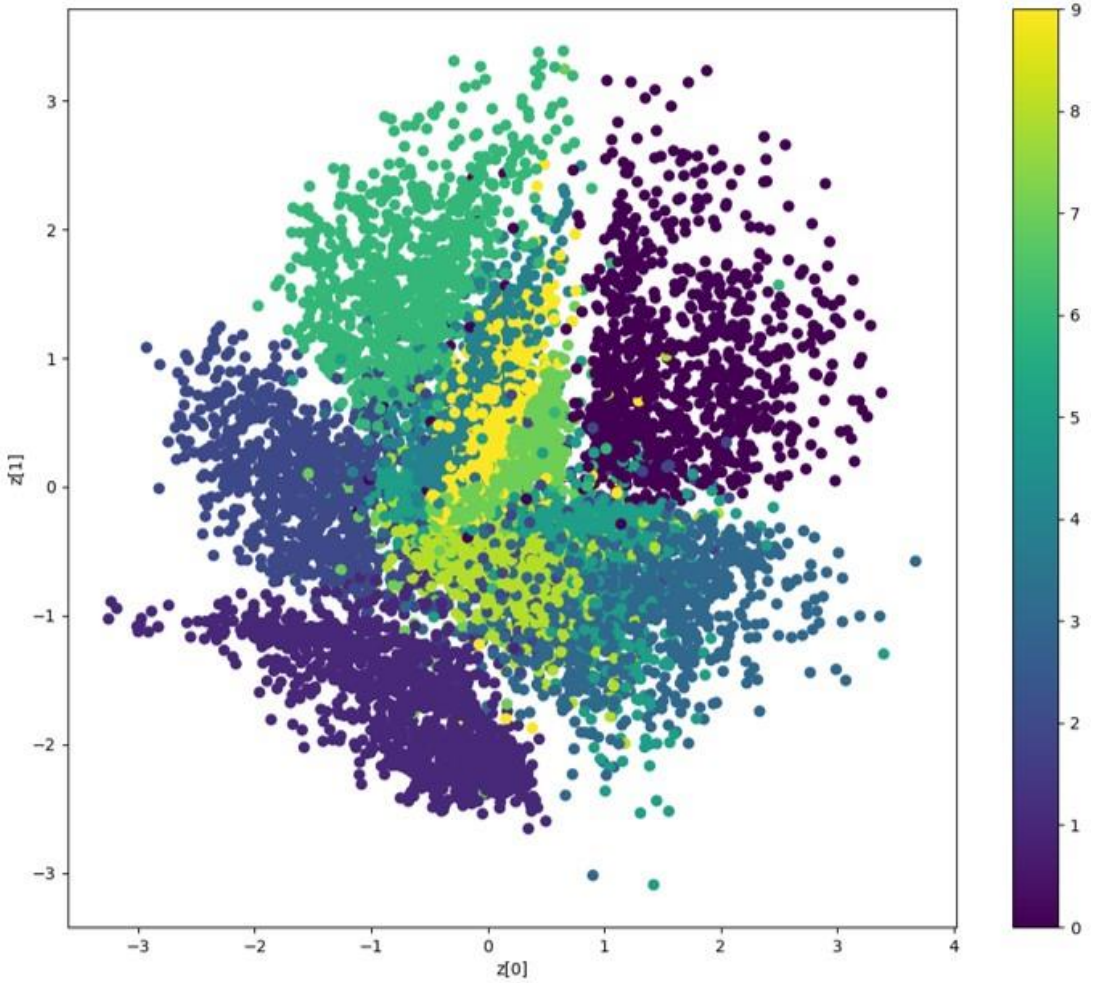


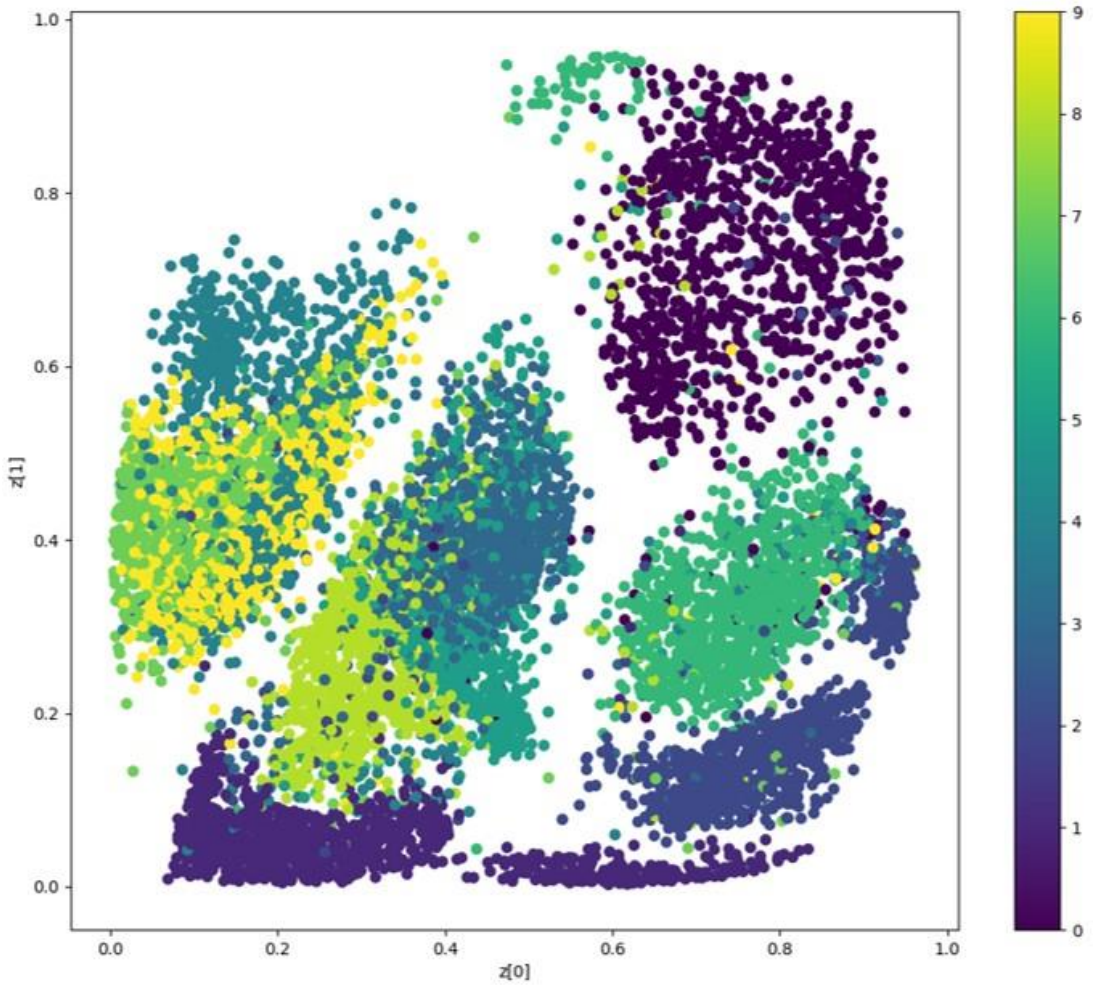
Fig credit : https://en.wikipedia.org/wiki/MNIST_database
https://peltarion.com/knowledge-center/documentation/datasets-view/datasets-used-in-tutorials/fashion-mnist-dataset

RESULTS

Basic VAE



B-Caps

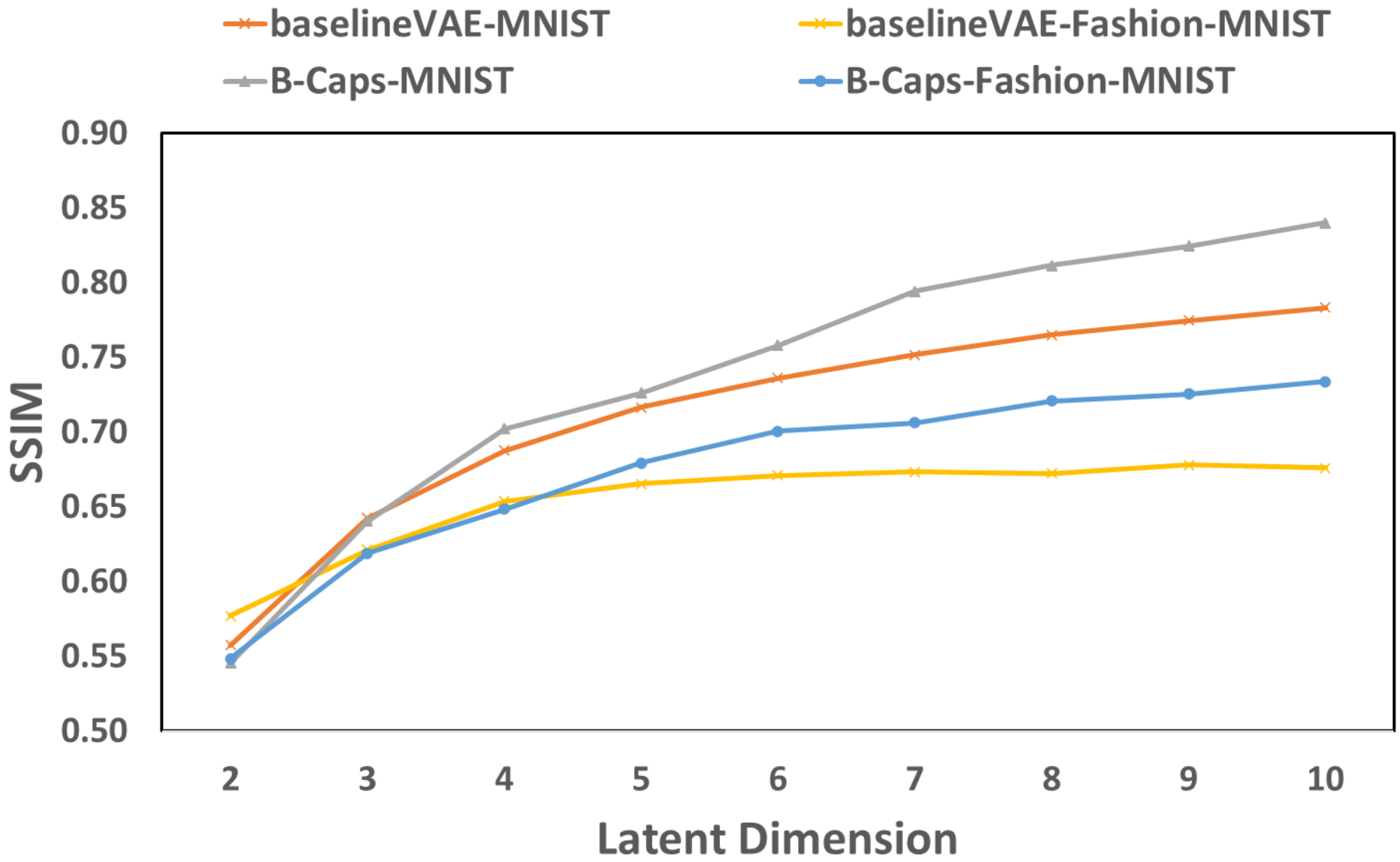


RESULTS

COMPARISON OF RECONSTRUCTION QUALITY ON MNIST WHILE VARYING THE CAPSULE TYPES (C) AND DESCRIPTION LENGTH (D). STD - STANDARD DEVIATION.

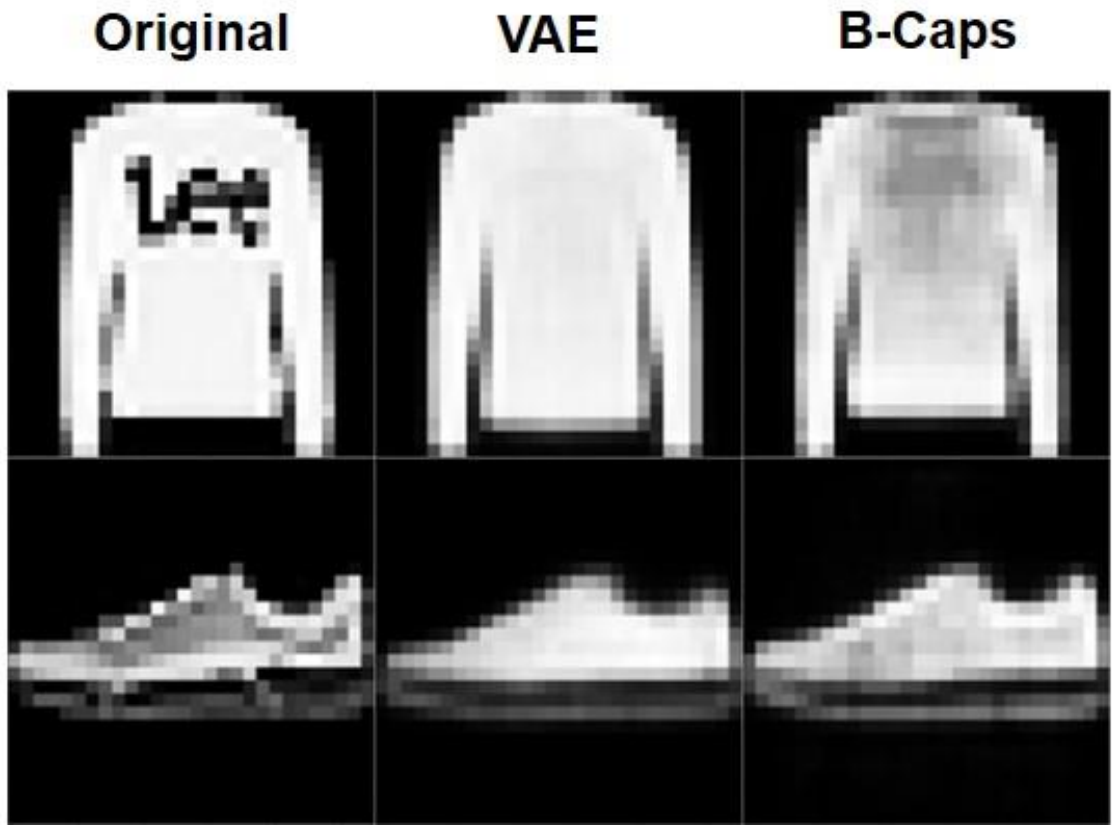
Model	Capsule types (C)	Description length (D)	SSIM mean±std ↑	MSE mean±std ↓
Baseline VAE	–	–	0.555 ± 0.154	0.041 ± 0.019
B-Caps	8	64	0.541 ± 0.144	0.043 ± 0.020
B-Caps	**16**	**64**	$\mathbf{0.580 \pm 0.133}$	$\mathbf{0.040 \pm 0.017}$
B-Caps	32	64	0.573 ± 0.147	0.041 ± 0.019
B-Caps	8	128	0.529 ± 0.152	0.046 ± 0.020
B-Caps	16	128	0.577 ± 0.129	0.040 ± 0.018

QUANTITATIVE RECONSTRUCTION MEASURE



QUALITATIVE RECONSTRUCTION

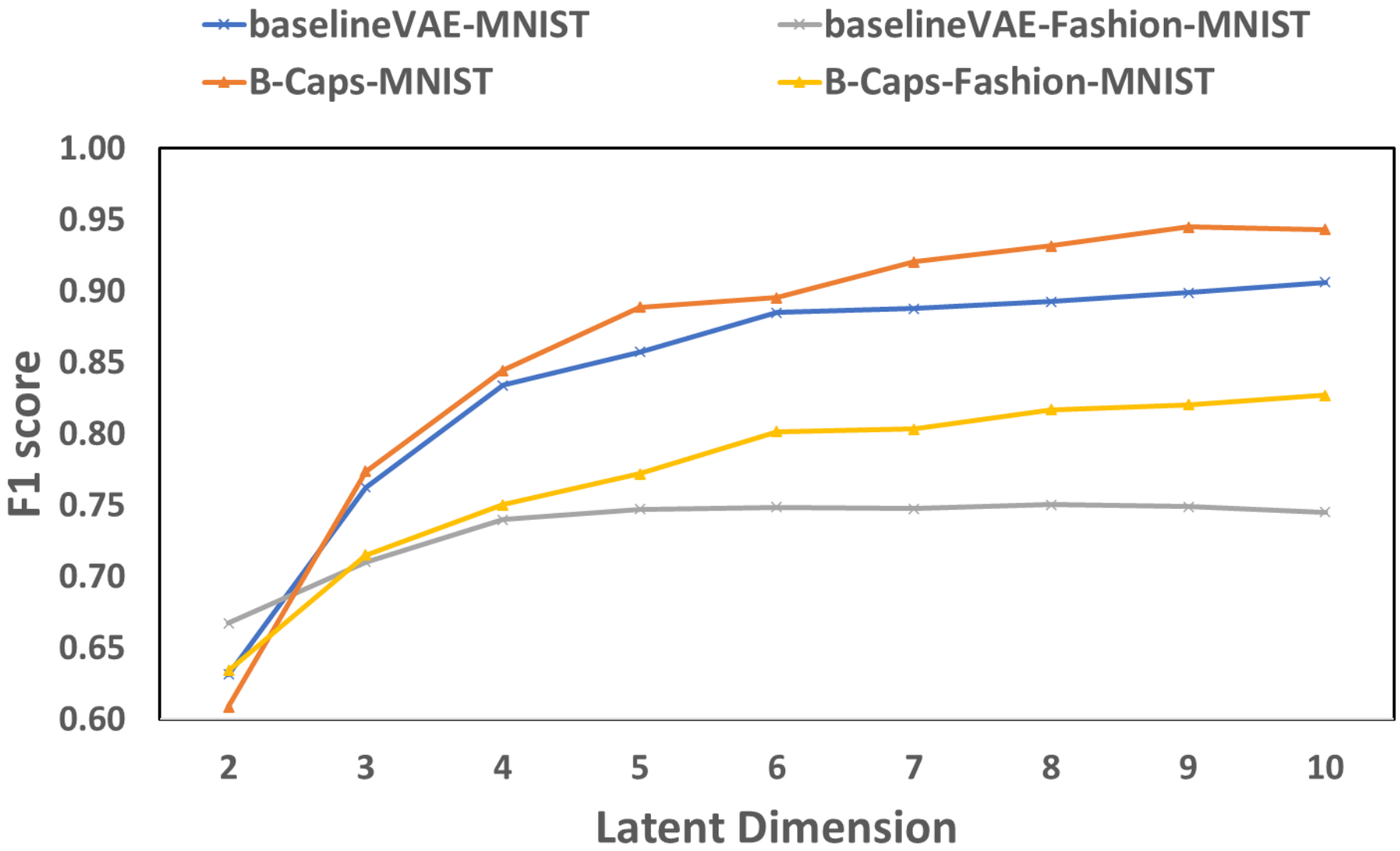
- Comparing the reconstruction quality of VAE vs B-Caps on Fashion-MNIST



CLASSIFICATION RESULTS

- Train SVM to classify reconstructed images

CLASSIFICATION RESULTS



SUMMARY

- Proposed novel **variational capsule encoder** which helps learn a potentially stronger latent space.
- Even in a shallow network, B-Caps can learn a more representative latent feature space compared to VAE.
- B-Caps has superiority in learning latent attributes in more complex datasets (FASHION-MNIST) compared to baseline VAE.

Thank you!

Variational Capsule Encoder

Poster #1595

Harish RaviPrakash

Syed Muhammad Anwar

Ulas Bagci