



# TextCaps : Handwritten Character Recognition with Very Small Datasets

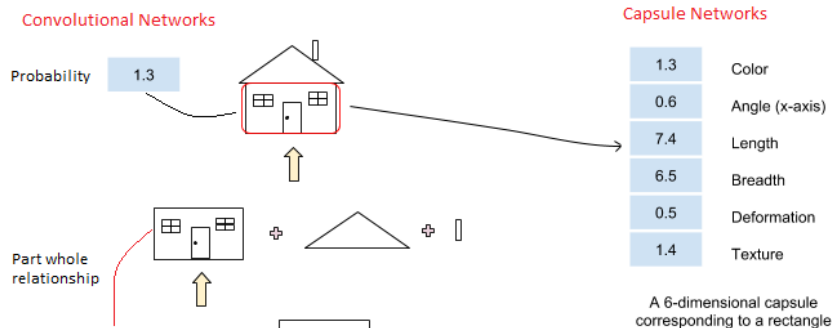
Vinoj Jayasundara, Sandaru Jayasekara, Hirunima Jayasekara, Jathushan Rajasegaran, Suranga Seneviratne\* and Ranga Rodrigo

University of Moratuwa,\* University of Sydney

WACV 2019  
January 08, 2019

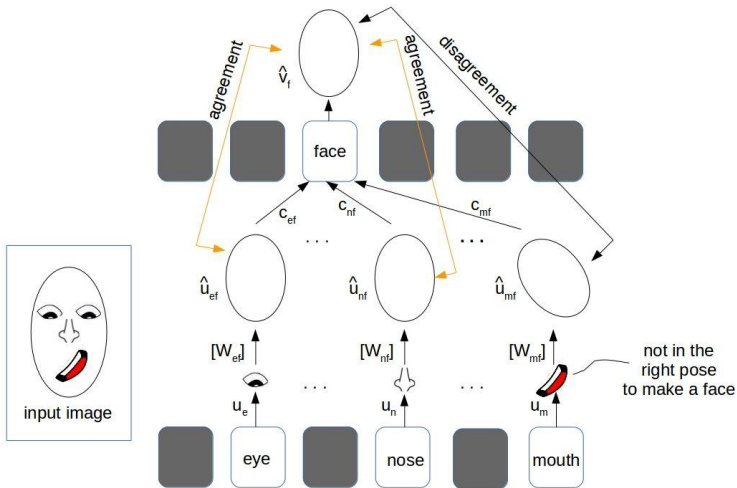
# Capsule Network : Instantiation parameters

Capsule Networks can encode any entity in **instantiation parameters**.



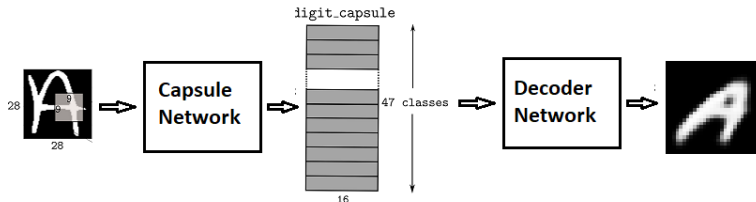
# Capsule Network : Routing by agreement

Capsule Networks propose a novel **routing by agreement algorithm**.



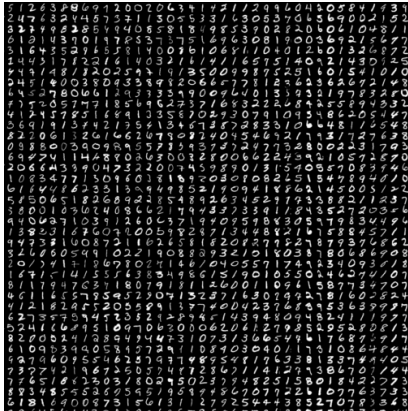
# Capsule Network : Decoder Network

Instantiation parameters can be used to **reconstruct** the entity back using a decoder network.



# Motivation

- Necessity of huge datasets for deep learning



- Localized languages cannot reap the benefits of deep learning due to the lack of sufficient data.

# Existing Solutions

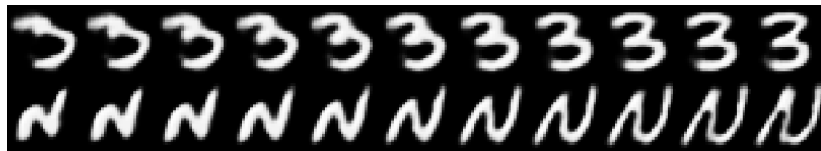
- Data augmentation (jittering, flipping): **Too Simple variations**
- GANs [1]: **Need a GAN per each class, still not effective**
- VAEs [2]: **High susceptibility of generating wrongly labelled data**

All of these are unable to attain realistic new data generation. Our Task - A novel data generation technique with high controllability, which can generate data with less cost, while producing realistic outputs.

# Existing Solutions

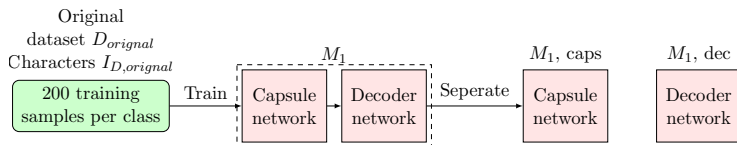
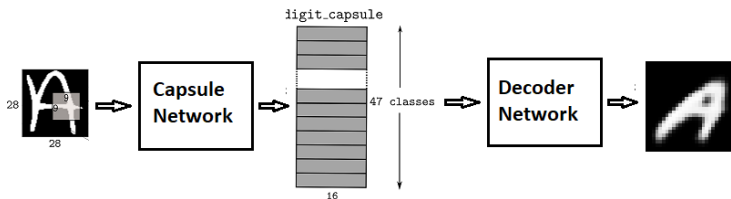
- Data augmentation (jittering, flipping): **Too Simple variations**
- GANs [1]: **Need a GAN per each class, still not effective**
- VAEs [2]: **High susceptibility of generating wrongly labelled data**

All of these are unable to attain realistic new data generation. Our Task - A novel data generation technique with high controllability, which can generate data with less cost, while producing realistic outputs.



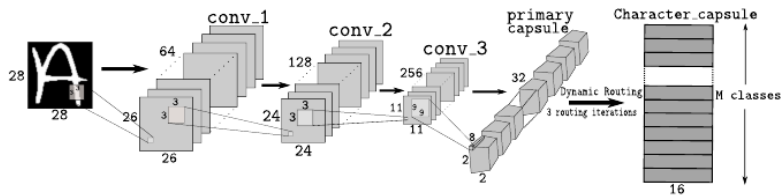
# Problem Definition

- Training with very small datasets (200 training samp/class)

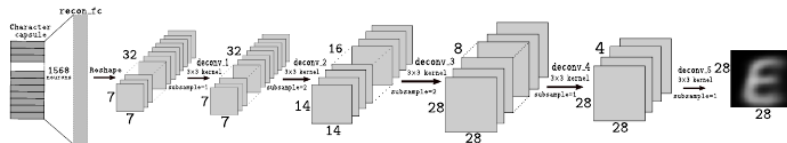




# Problem Definition



Capsule Network



Decoder Network

# Problem Definition

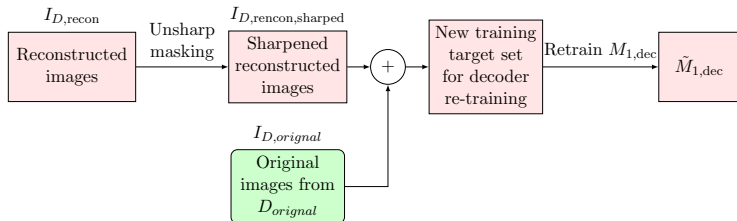
- Results in,



- Two main issues identified
  - ① The reconstructed images are blurry
  - ② The subtle variations in the characters are not properly captured

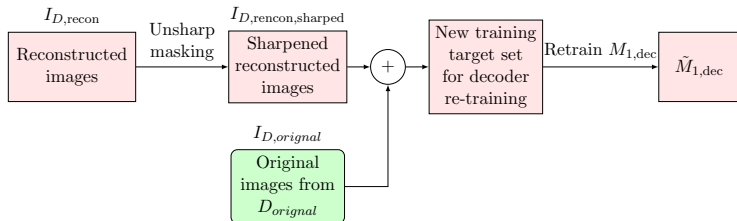
# Decoder Re-training Technique

- Decoder Re-training



# Decoder Re-training Technique

- Decoder Re-training



- Results in,

Original image

Reconstructed image

After decoder re-training



# Data Generation Technique

- Perturbation of instantiation Parameters can generate human-like variations



# Data Generation Technique

- Perturbation of instantiation Parameters can generate human-like variations



- Uncontrolled perturbation can cause distortions
  - ① Visually unrecognizable images



# Data Generation Technique

- Perturbation of instantiation Parameters can generate human-like variations



- Uncontrolled perturbation can cause distortions
  - 1 Visually unrecognizable images

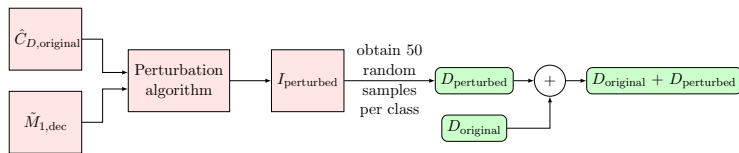


- 2 Class jumps



# Data Generation Technique

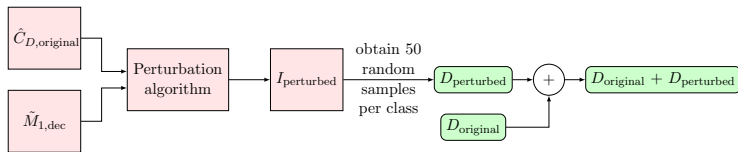
- New image data generation





# Data Generation Technique

- New image data generation



- Results in,

Original image

---

Trained with original dataset

---

Trained with generated dataset



# Performance

- We use five benchmark datasets to evaluate our system

EMNIST-Letters		
Implementation	With full train set	With 200 samp/class
Cohen <i>et al.</i> [1]	85.15%	-
Wiyatnoet <i>al.</i> [2]	91.27%	-
<b>TextCaps</b>	<b>95.36 ± 0.30%</b>	<b>92.79 ± 0.30%</b>

EMNIST-Balanced		
Implementation	With full train set	With 200 samp/class
Cohen <i>et al.</i> [1]	78.02%	-
Dufourq <i>et al.</i> [3]	88.3%	-
<b>TextCaps</b>	<b>90.46 ± 0.22%</b>	87.82 ± 0.25%

EMNIST-Digits		
Implementation	With full train set	With 200 samp/class
Cohen <i>et al.</i> [1]	95.90%	-
Dufourq <i>et al.</i> [3]	99.3%	-
<b>TextCaps</b>	<b>99.79 ± 0.11%</b>	98.96 ± 0.22%

# Performance

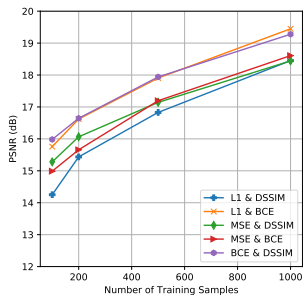
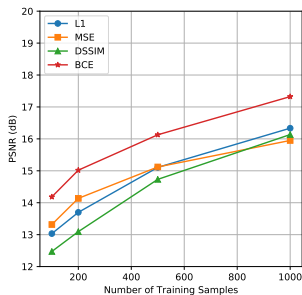
<b>MNIST</b>		
Implementation	With full train set	With 200 samp/class
Sabour <i>et al.</i> [4]	99.75%	-
Cireřan <i>et al.</i> [5]	99.77%	-
Wan <i>et al.</i> [6]	<b>99.79%</b>	-
<b>TextCaps</b>	99.71 $\pm$ 0.18%	98.68 $\pm$ 0.30%

<b>Fashion MNIST</b>		
Implementation	With full train set	With 200 samp/class
Xiao <i>et al.</i> [7]	89.7%	-
Bhatnagar <i>et al.</i> [8]	92.54%	-
Zhong <i>et al.</i> [9]	<b>96.35%</b>	-
<b>TextCaps</b>	93.71 $\pm$ 0.64%	85.36 $\pm$ 0.79%

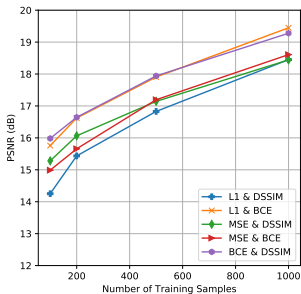
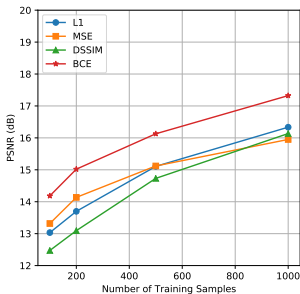
# Loss Function Analysis

- Loss function of the decoder has a direct impact on the Reconstruction performance



# Loss Function Analysis

- Loss function of the decoder has a direct impact on the Reconstruction performance



Original

DSSIM

BCE

BCE & DSSIM



# In Conclusion

In TextCaps, we present a novel system which consists of decoder re-training and data generation techniques, which creates






- Images more realistic than existing techniques
- Starting from a very low amount of data
- Generate images as much as necessary
- Without any user interaction or post-processing

Thank you!





# Thank You!

Paper ID : 535

Poster Session : Tuesday 19:30 - 22:00 @ Kona 1-3

-  Cohen, G., Afshar, S., Tapson, J., van Schaik, A.:  
EMNIST: an extension of MNIST to handwritten letters.  
CoRR (2017)
-  Wiyatno, R., Orchard, J.:  
Style memory: Making a classifier network generative.  
CoRR (2018)
-  Dufourq, E., Bassett, B.A.:  
Eden: Evolutionary deep networks for efficient machine  
learning.  
In: PRASA-RobMech, Bloemfontein, South Africa (2017)  
110–115
-  Sabour, S., Frosst, N., Hinton, G.E.:  
Dynamic routing between capsules.  
In: NIPS, Long Beach, CA (2017) 3856–3866
-  Ciresan, D.C., Meier, U., Schmidhuber, J.:  
Multi-column deep neural networks for image classification.  
CoRR (2012)



-  Wan, L., Zeiler, M., Zhang, S., Cun, Y.L., Fergus, R.:  
Regularization of neural networks using dropconnect.  
ICML **28** (2013) 1058–1066
-  Xiao, H., Rasul, K., Vollgraf, R.:  
Fashion-mnist: a novel image dataset for benchmarking  
machine learning algorithms.  
CoRR (2017)
-  Bhatnagar, S., Ghosal, D., Kolekar, M.H.:  
Classification of fashion article images using convolutional  
neural networks.  
In: ICIP, Shimla, India (2017) 1–6
-  Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.:  
Random erasing data augmentation.  
CoRR (2017)