A Deep Learning Approach to Sign Language Recognition using Stacked Sparse Autoencoders

Dr. Pablo Rivas*

Ezequiel R.^ Omar V.^ Samuel G^ Deep D.*

* *Marist College,* School of Computer Science and Mathematics

^ *Nogales Institute of Technology*, Research and Grad. Studies

MARIST

The Problem

- American Sign Language (ASL)

MARIST

The Data

- ***Learning*** the American Sign Language (ASL)

MARIST

The Data

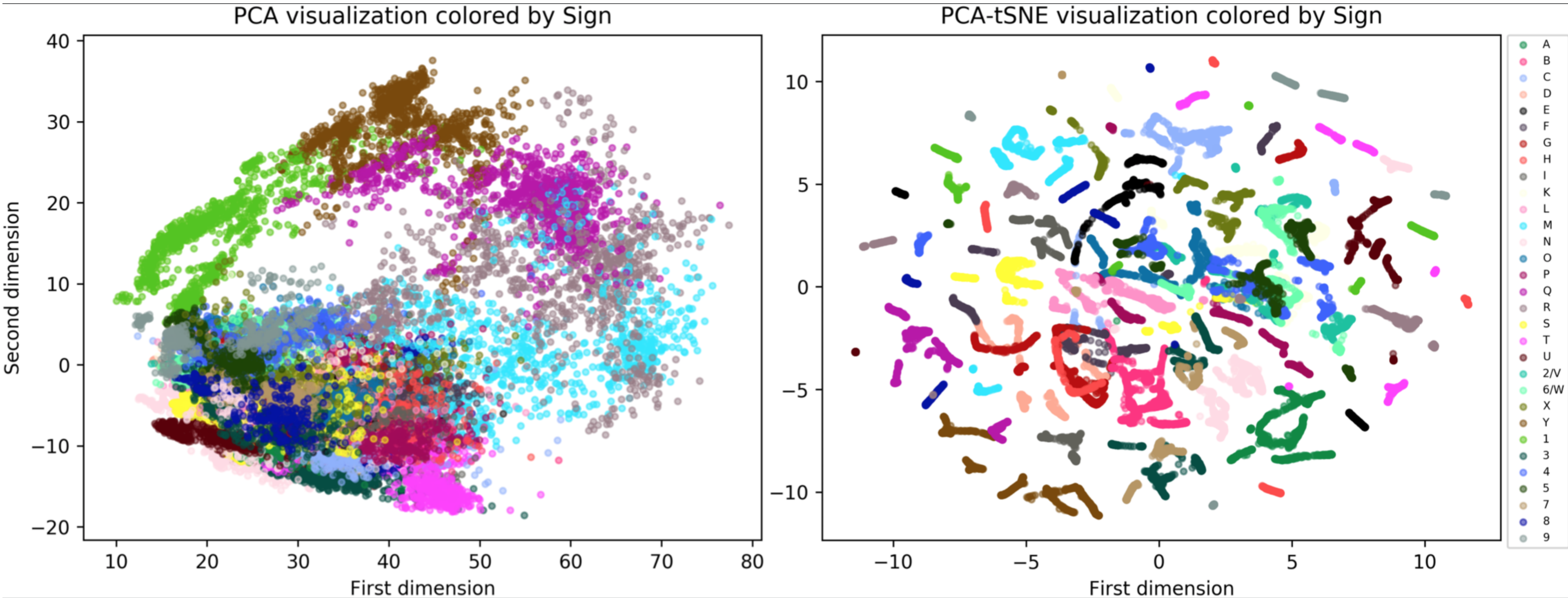
- ***Learning*** the American Sign Language (ASL)

The Data

- ***Learning*** the American Sign Language (ASL)

MARIST

The Data



MARIST

Existing Approaches

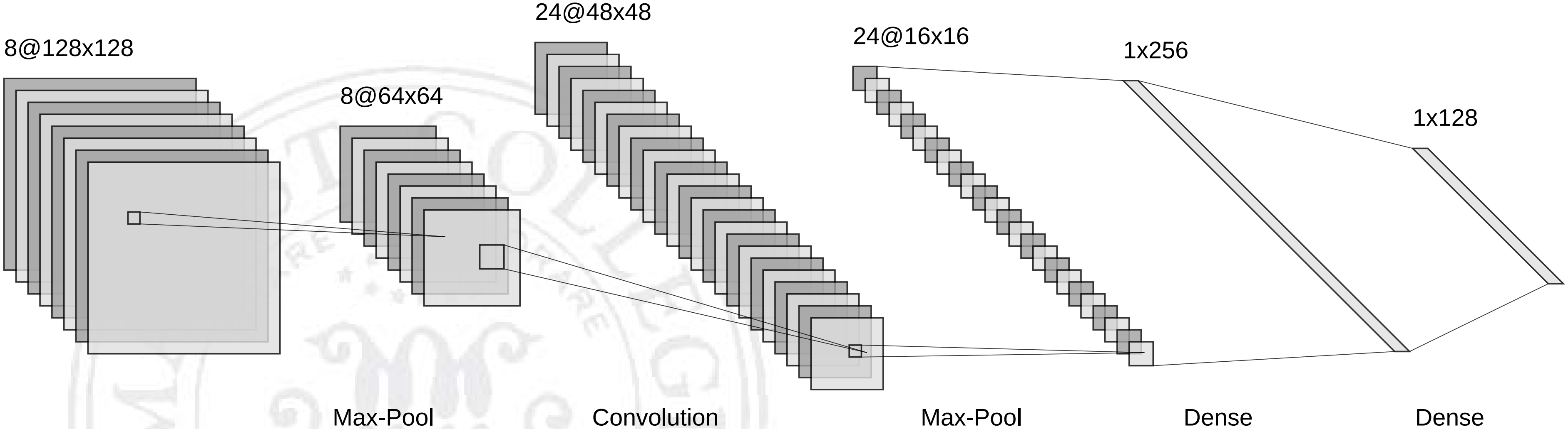
Method	Class type	# of class	# of subj.	Test w/ diff.	Input	Accur.(%)
Nagi *et al.* [8]	Gesture	6	-	No	Color	96
Van den Bergh *et al.* [14]	Gesture	6	-	No	Color & Depth	99.54
Isaacs *et al.* [3]	Alphabets	24	-	-	Color	99.9
Pugeault *et al.* [10]	Alphabets	24	5	-	Color	73
Pugeault *et al.* [10]	Alphabets	24	5	-	Depth	69
Pugeault *et al.* [10]	Alphabets	24	5	-	Color & Depth	75
Kuznetsova *et al.* [6] (50/50)%	Alphabets	24	5	No	Depth	87
Kuznetsova *et al.* [6] (4/1)	Alphabets	24	5	Yes	Depth	57
Dong *et al.* [2] (50/50)%	Alphabets	24	5	No	Depth	90
Dong *et al.* [2] (4/1)	Alphabets	24	5	Yes	Depth	70
Ours (re-training) (50/25/25)%	Alph. & Digit	31	5	No	Depth	99.99
Ours (re-training) (3/1/1)	Alph. & Digit	31	5	Yes	Depth	75.18
Ours (re-training) (4/1)	Alph. & Digit	31	5	Yes	Depth	78.39
Ours (fine-tuning) (3/1/1)	Alph. & Digit	31	5	Yes	Depth	83.58
Ours (fine-tuning) (4/1)	Alph. & Digit	31	5	Yes	Depth	85.49

- Kang, B., Tripathi, S., Nguyen, T.Q.: Real-time sign language fingerspelling recognition using convolutional neural networks from depth map. In: Pattern Recognition (ACPR), 2015 3rd IAPR Asian Conference on, pp. 136–140. IEEE (2015)

MARIST

Existing Solutions

- Learning the American Sign Language (ASL) ***with CNNs***



MARIST

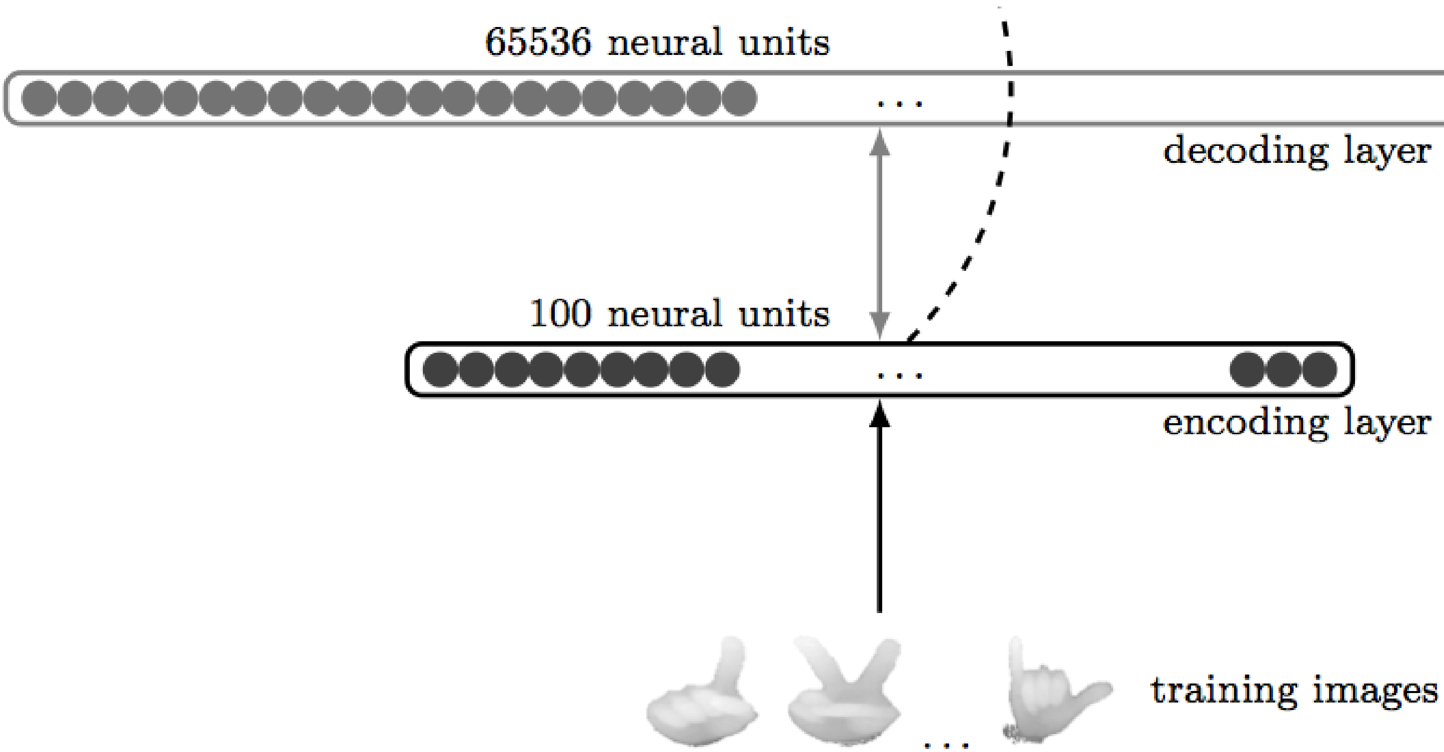
The Proposal

- Learning the American Sign Language (ASL) ***with Auto-encoders***
  - ***Simpler than CNN***
  - ***More efficient than CNN (deployed)***
  - ***Faster to train than CNN (for a similar number of layers)***
  - ***Similar performance to a CNN***
  - ***CNNs are not the panacea in pattern recognition on images or computer vision***

MARIST

The Proposal

- ***Lets talk about Auto-encoders***



MARIST

The Proposal

- ***Lets talk about Auto-encoders***

$$L = \frac{1}{N} \left\| \mathbf{x}_n - \hat{\mathbf{x}}_n \right\|_2^2 + \theta_w \frac{1}{2} \sum_{l=1}^{L} \left\| \mathbf{w}^l \right\|_2^2 + \theta_s \sum_{m=1}^{M} KL \left(\theta_\alpha \left\| \bar{\alpha}_m \right. \right)$$

MARIST

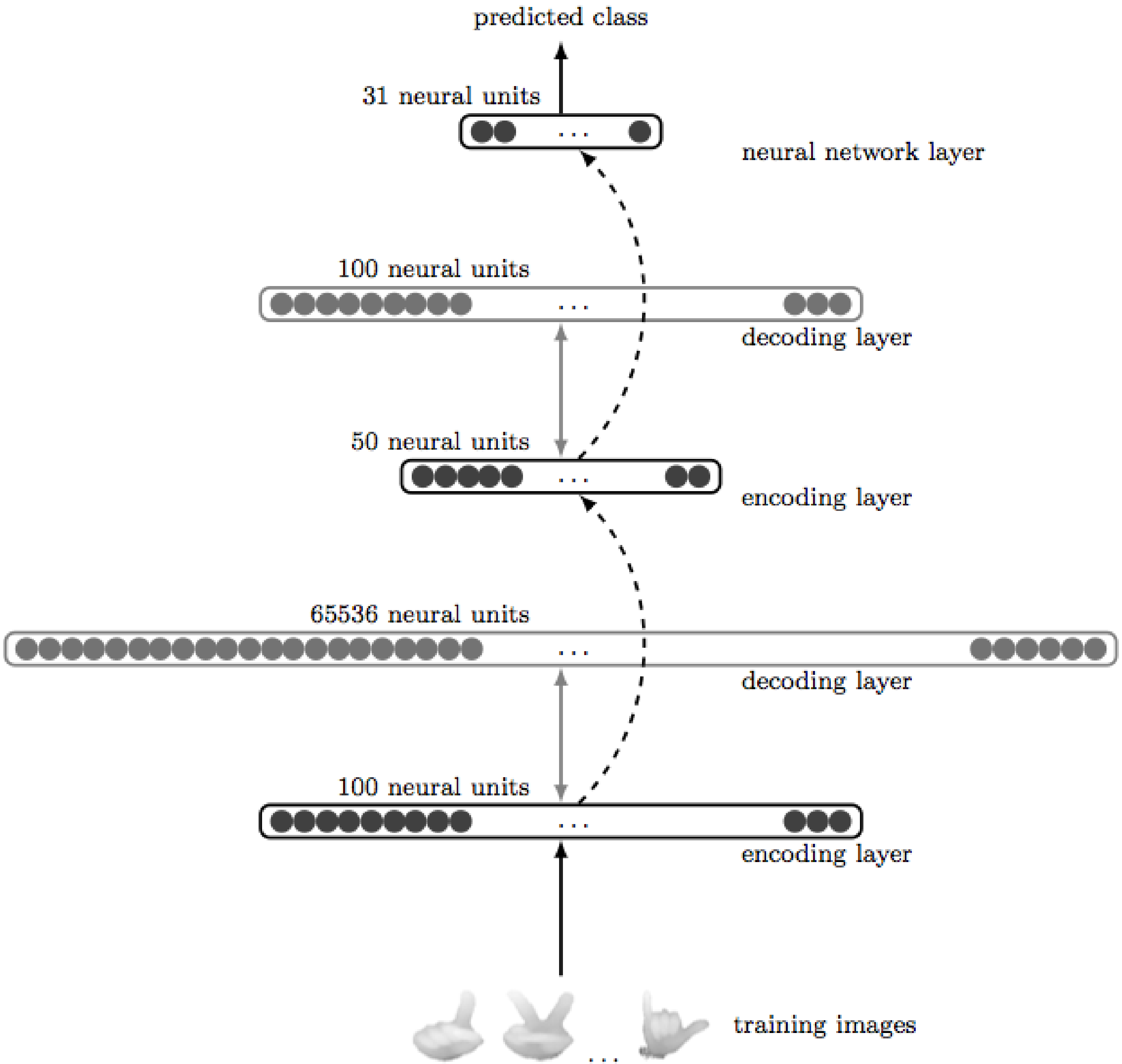
The Proposal

- ***Lets talk about Auto-encoders***

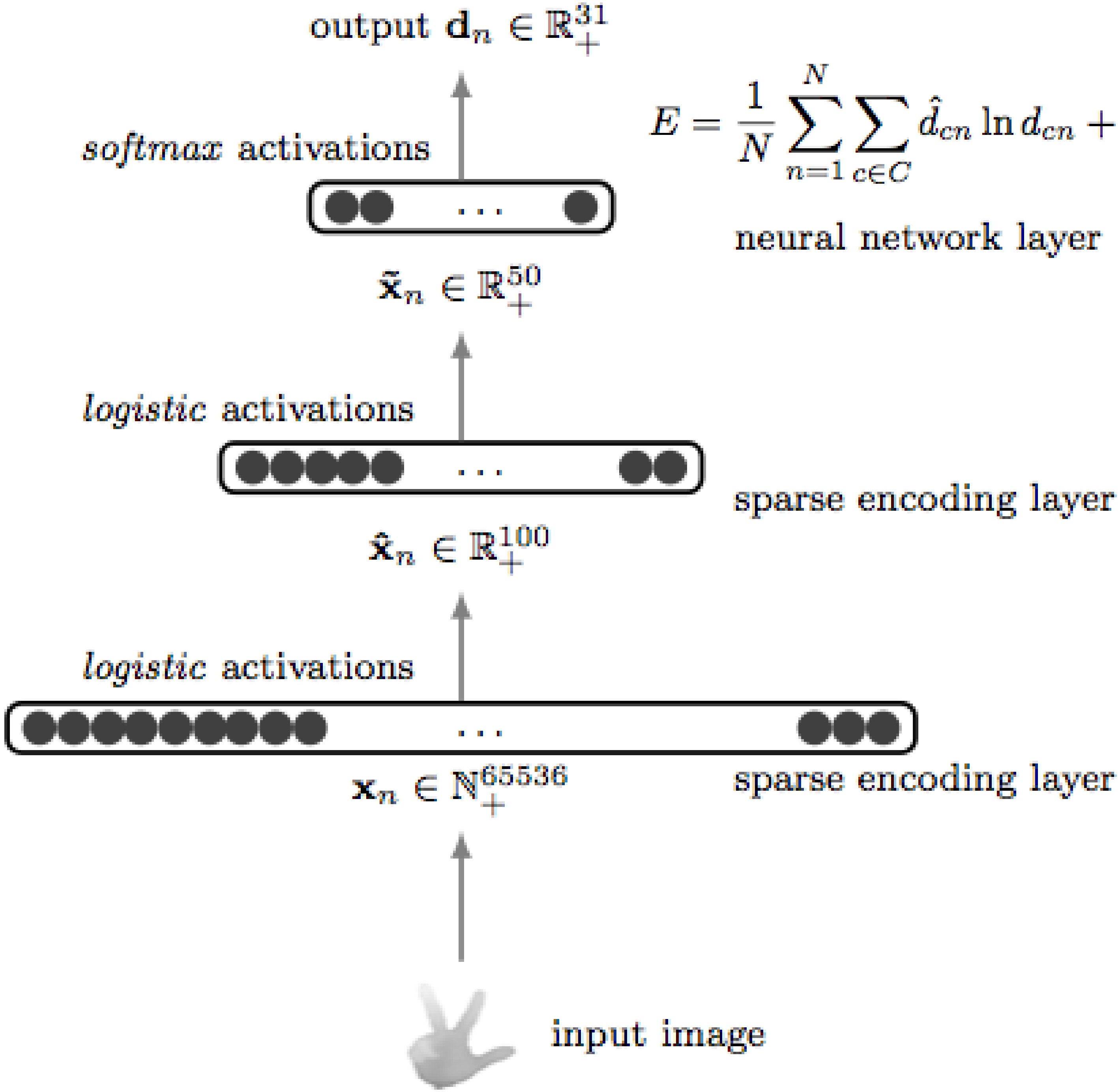
$$L = \frac{1}{N}\left\|\mathbf{x}_n - \hat{\mathbf{x}}_n\right\|_2^2 + \theta_w \frac{1}{2}\sum_{l=1}^{L}\left\|\mathbf{w}^l\right\|_2^2 + \theta_s \sum_{m=1}^{M} KL\left(\theta_\alpha \| \bar{\alpha}_m\right)$$

$$\sum_{m=1}^{M} KL\left(\theta_\alpha \| \bar{\alpha}_m\right) = \sum_{m=1}^{M} \theta_\alpha \log\left(\frac{\theta_\alpha}{\bar{\alpha}_m}\right) + (1-\theta_\alpha)\log\left(\frac{1-\theta_\alpha}{1-\bar{\alpha}_m}\right)$$

$$\bar{\alpha}_m = \frac{1}{N}\sum_{n=1}^{N}\psi\left(\mathbf{w}_m^{(l)T}\mathbf{x}_n + b_m^{(l)}\right)$$



predicted class
31 neural units
neural network layer
100 neural units
decoding layer
50 neural units
encoding layer
65536 neural units
decoding layer
100 neural units
encoding layer
training images



$$E = \frac{1}{N} \sum_{n=1}^{N} \sum_{c \in C} \hat{d}_{cn} \ln d_{cn} + (1 - \hat{d}_{cn}) \ln(1 - d_{cn})$$

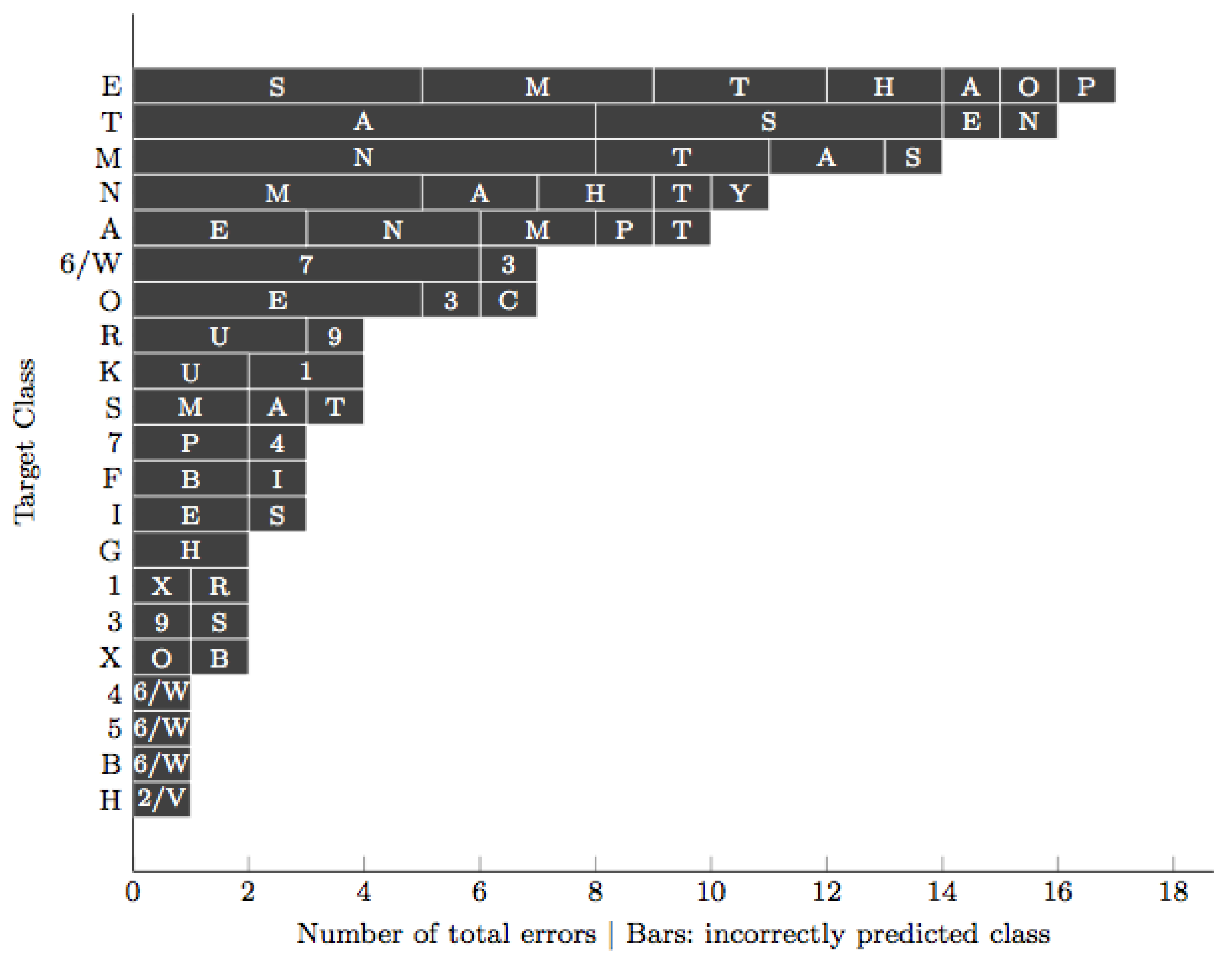
Results

- ***of the Auto-encoders***

	S1	S2	S3	S4	S5	Avg.
ACC	0.9748	0.9923	0.9935	0.9929	0.9910	0.9889
SPC	0.9991	0.9997	0.9998	0.9998	0.9997	0.9996
MAE	0.1483	0.0640	0.0373	0.0347	0.0494	0.0667

MARIST

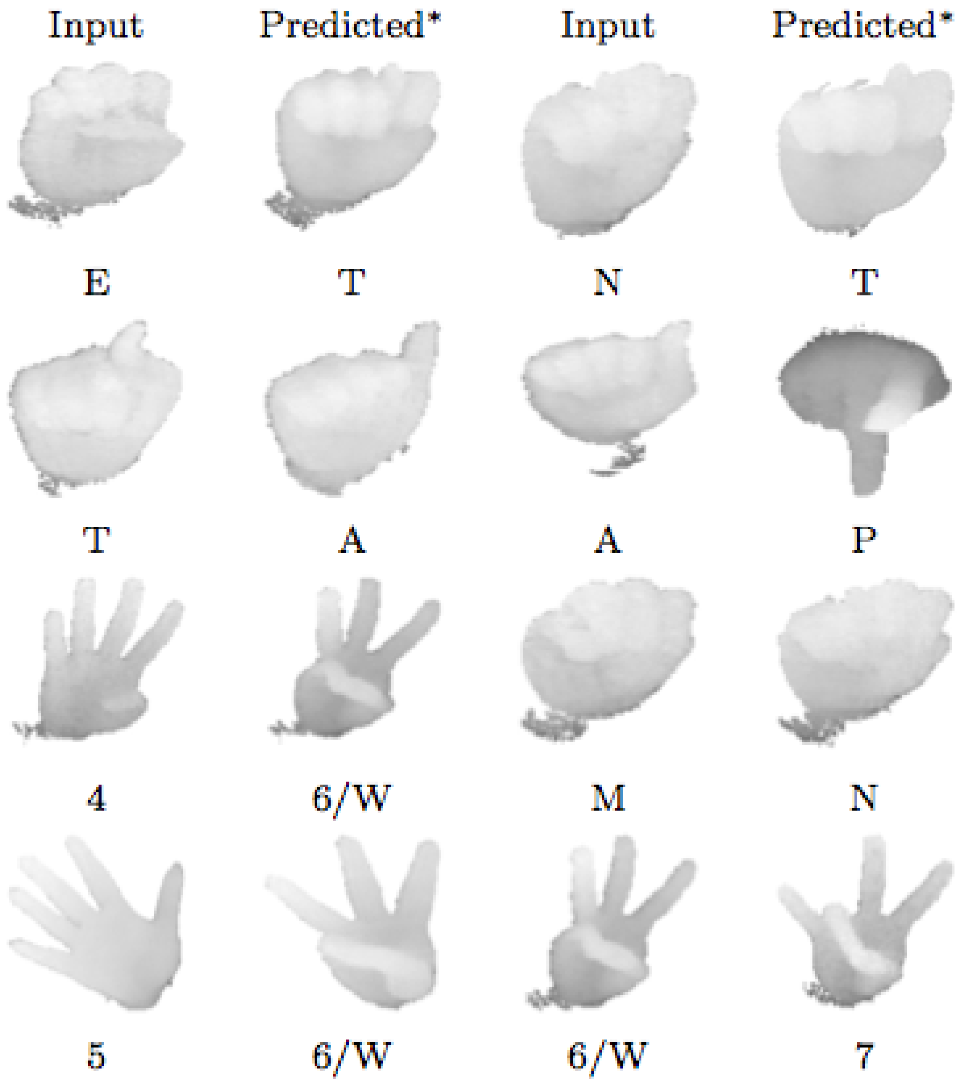
Results

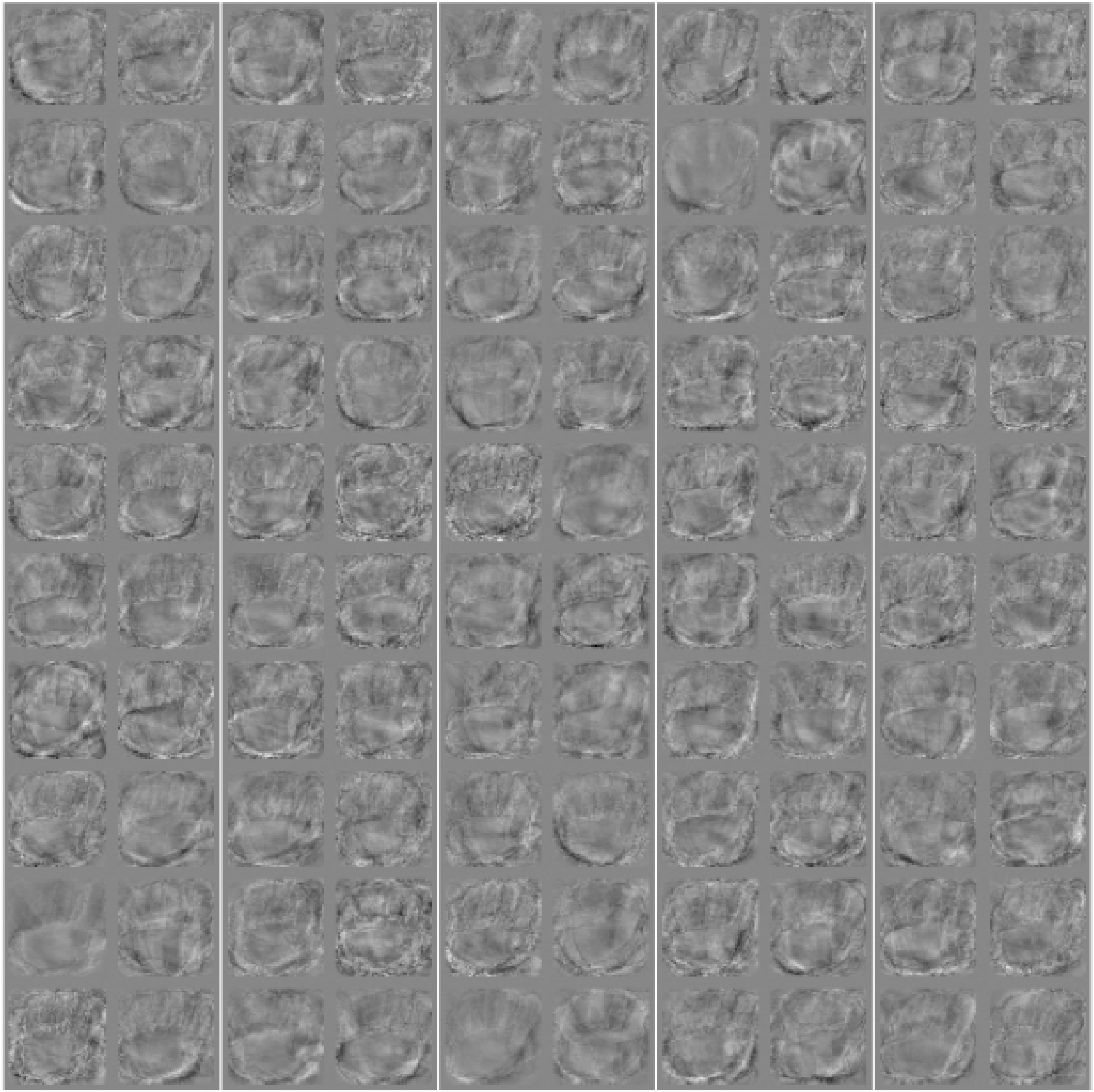


Target Class
E
T
M
N
A
6/W
O
R
K
S
7
F
I
G
1
3
X
4
5
B
H
S M T H A O P
A S E N
N T A S
M A H T Y
E N M P T
7 3
E 3 C
U 9
U 1
M A T
P 4
B I
E S
H
X R
9 S
O B
6/W
6/W
6/W
2/V
0 2 4 6 8 10 12 14 16 18
Number of total errors | Bars: incorrectly predicted class

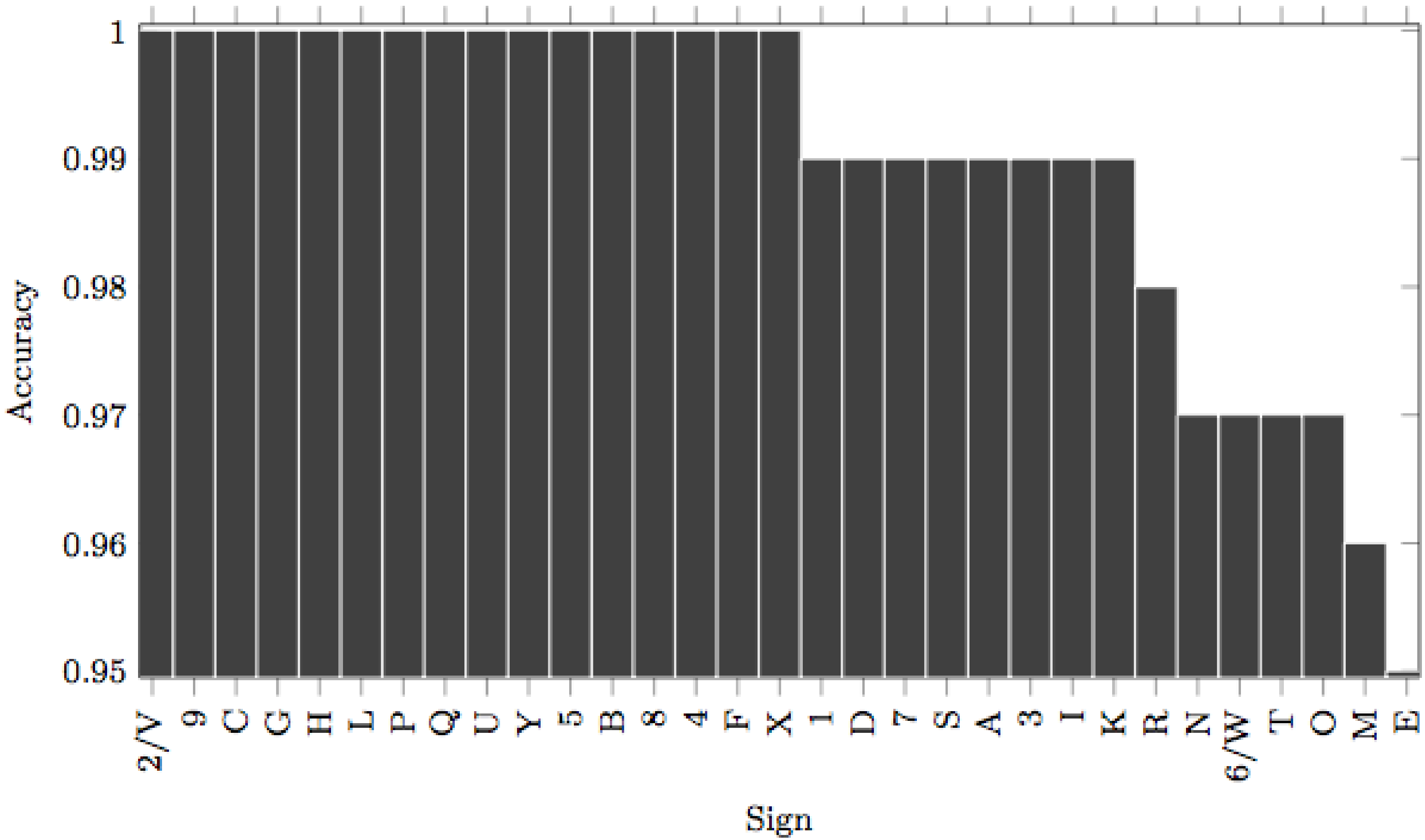

A B C D E F G H I J K L M N O P Q
R S T U V W X Y Z 1 2 3 4 5 6 7 8 9

Results



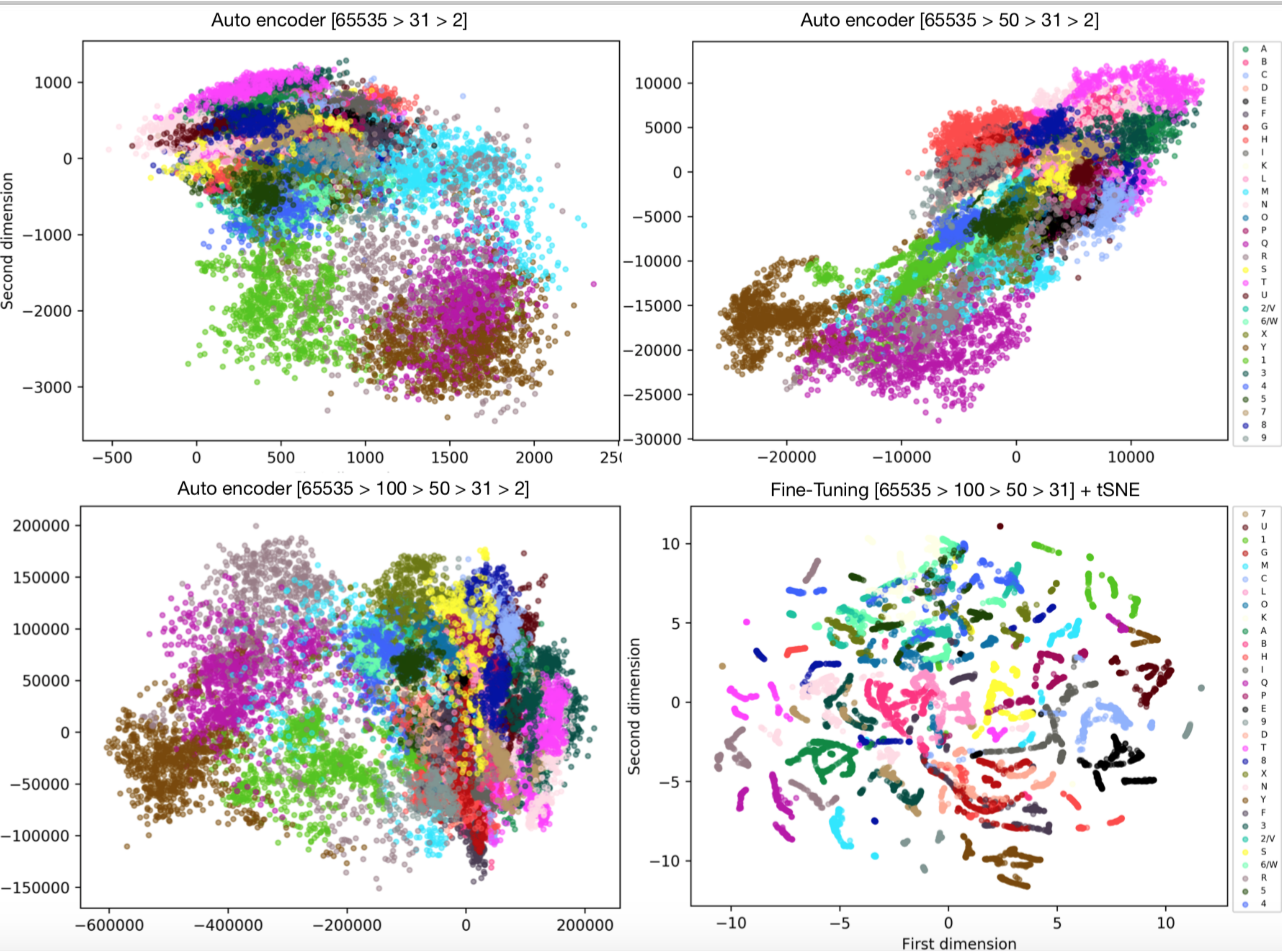


Results



Accuracy
1
0.99
0.98
0.97
0.96
0.95
2/V 9 C G H L P Q U Y 5 B 8 4 F X 1 D 7 S A 3 I K R N 6/W T O M E
Sign

MARIST



Auto encoder [65535 > 31 > 2]
Auto encoder [65535 > 50 > 31 > 2]
Auto encoder [65535 > 100 > 50 > 31 > 2]
Fine-Tuning [65535 > 100 > 50 > 31] + tSNE
Second dimension
First dimension
A
B
C
D
E
F
G
H
I
K
L
M
N
O
P
Q
R
S
T
U
2/V
6/W
X
Y
1
3
4
5
7
8
9

Chameleon Setup

- ***Resource type:*** bare metal/CHI@TACC
- ***Lease:*** GPU P100
- ***Image:*** CC-Ubuntu16.04-CUDA8
- ***Libraries:***
  - **`cuDNN`**
  - **`libatlas-dev`**

MARIST

Libraries: `cuDNN`

cc@fox: ~ — ssh — 90×20 — ⌘1

```
Connection closed by 129.114.109.140
Pablos-MacBook-Air:.ssh rivas_perea$ ssh cc@129.114.109.140
Welcome to Ubuntu 16.04.2 LTS (GNU/Linux 4.4.0-72-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

  Get cloud support with Ubuntu Advantage Cloud Guest:
    http://www.ubuntu.com/business/services/cloud

0 packages can be updated.
0 updates are security updates.


Last login: Tue Sep 12 16:35:00 2017 from 204.210.149.122
cc@fox:~$ sudo cp cuda/include/cudnn.h /usr/local/cuda/include
cc@fox:~$ sudo cp cuda/lib64/libcudnn* /usr/local/cuda/lib64
cc@fox:~$ sudo chmod a+r /usr/local/cuda/include/cudnn.h /usr/local/cuda/lib64/libcudnn*
cc@fox:~$
```

Chameleon Setup

- ***Resource type:*** bare metal/CHI@TACC
- ***Lease:*** GPU P100
- ***Image:*** CC-Ubuntu16.04-CUDA8
- ***Packages:***
  - `gcc, gfortran`
  - `python-{numpy scipy matplotlib}`
  - `tensorflow`
  - `glances, nvidia-ml-py, screen`

Conclusions

- Learning the American Sign Language (ASL) ***with Auto-encoders***
  - ***Simpler than CNN***
  - ***More efficient than CNN***
  - ***Faster to train than CNN***
  - ***Similar performance to a CNN***
  - ***CNNs are not the panacea in pattern recognition on images or computer vision*** *(no free lunch theorem)*

MARIST

A Deep Learning Approach to Sign Language Recognition using Stacked Sparse Autoencoders

Interested in code? Check Deep's repo:

https://github.com/DeepDand/research

MARIST