Comparaison of Neuronal Attention Models

Mohamed Karim BELAID
University of Passau, Germany
04.07.2019

1. Motivation

Problem: Street View House Number

Find the street house number using a high dimensional image

Figure 1: Examples of cropped images from the SVHN database [14].

Challenge Solved using CNN !

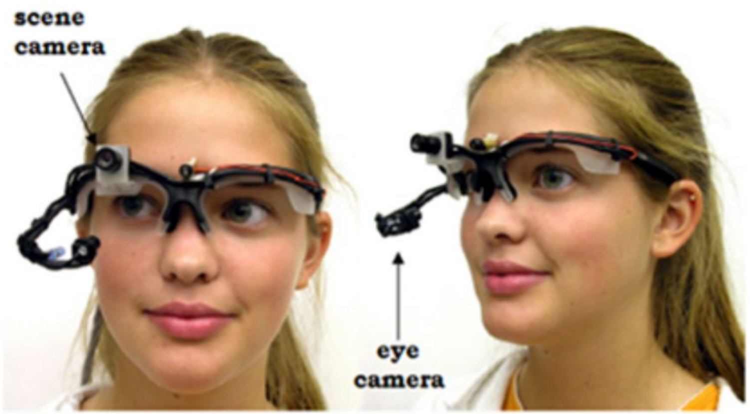
Functional requirement: accuracy 95% [4] ✔️

Non-functional requirement: 6 days of training ❗

→ Better Solution: **Visual Attention Model**

Eye movements in natural behavior

What controls attention in natural environments?



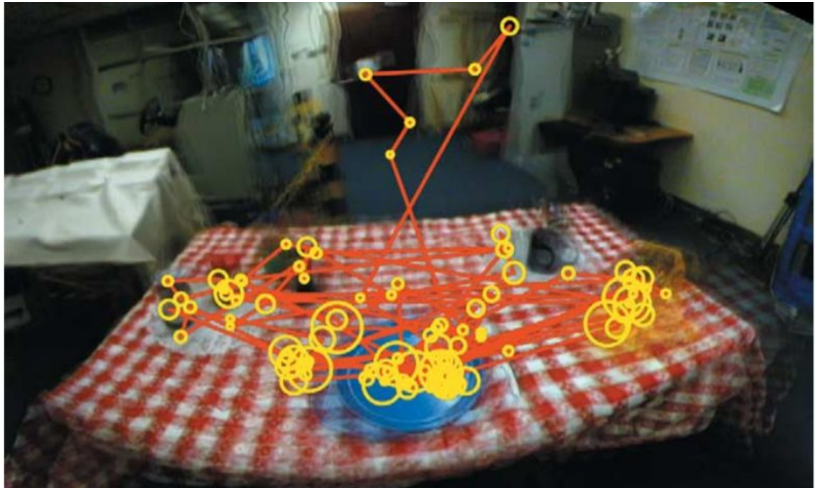


Figure 2: Portable, head mounted eye-tracker [5], composed of an infrared video camera (to follow the eye movement). The "scene camera" records the environment from the observer's viewpoint. Eye position is then superimposed on the video record.

Eye Movement model in 3 rules [5]

Rule 1

Sequential Analysis:

The eye analyses small regions of the environment in a sequential way

Rule 2

Short Term Memory:

The next target depends on the content of the current one

Rule 3

Internal Reward:

The results of this active search are transformed into a reward. $\Rightarrow$ the next picked step maximize the potential reward.

Eye Movement model in Python

Sequential Analysis → **For Loop**

Short Term Memory → **Recurrent Neural Net.**

Internal Reward → **Reinforcement Learning**

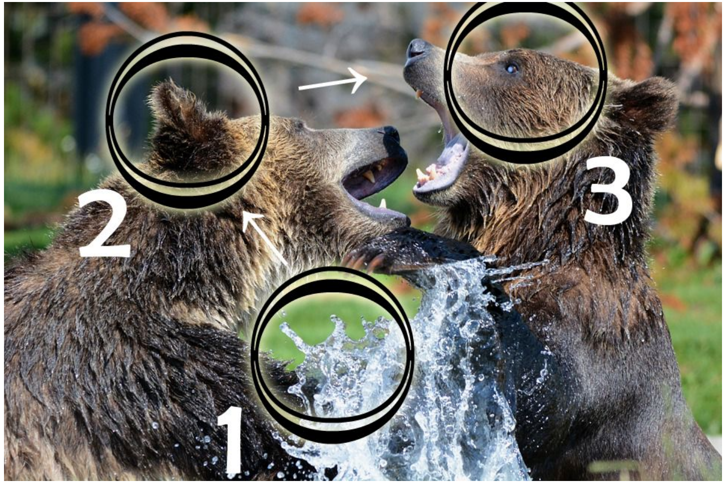


Figure 3: “2 Grizzly bears playing in water”

Agenda

1. Motivation
2. Neuronal Attention Model
3. Hyperparameters
4. Experiment
    - The MNIST dataset
    - Experiment setup
5. Results
6. Conclusion

2. Neuronal Attention Model

Glimpse sensor

Figure 4: Glimpse Sensor - Flow chart

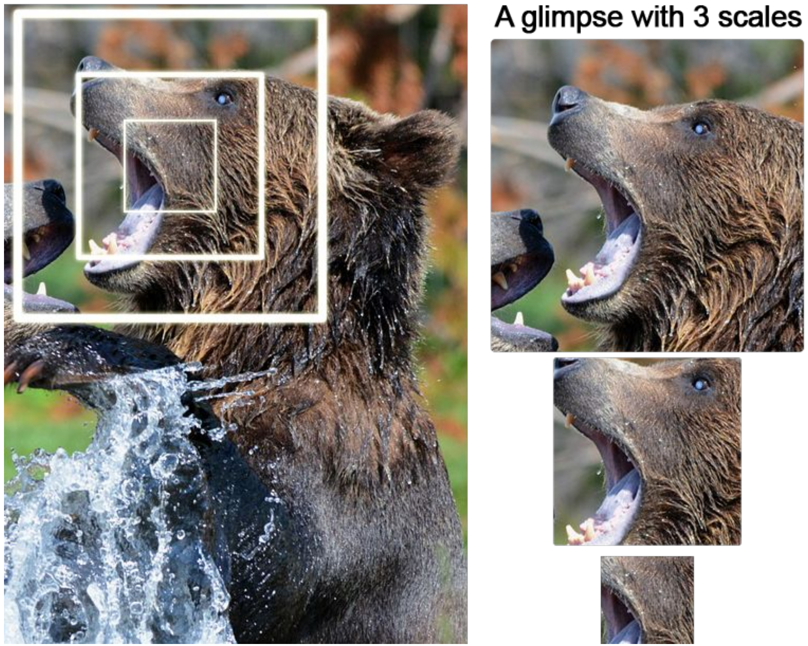


Figure 5: “2 Grizzly bears playing in water”

Neuronal Attention Model

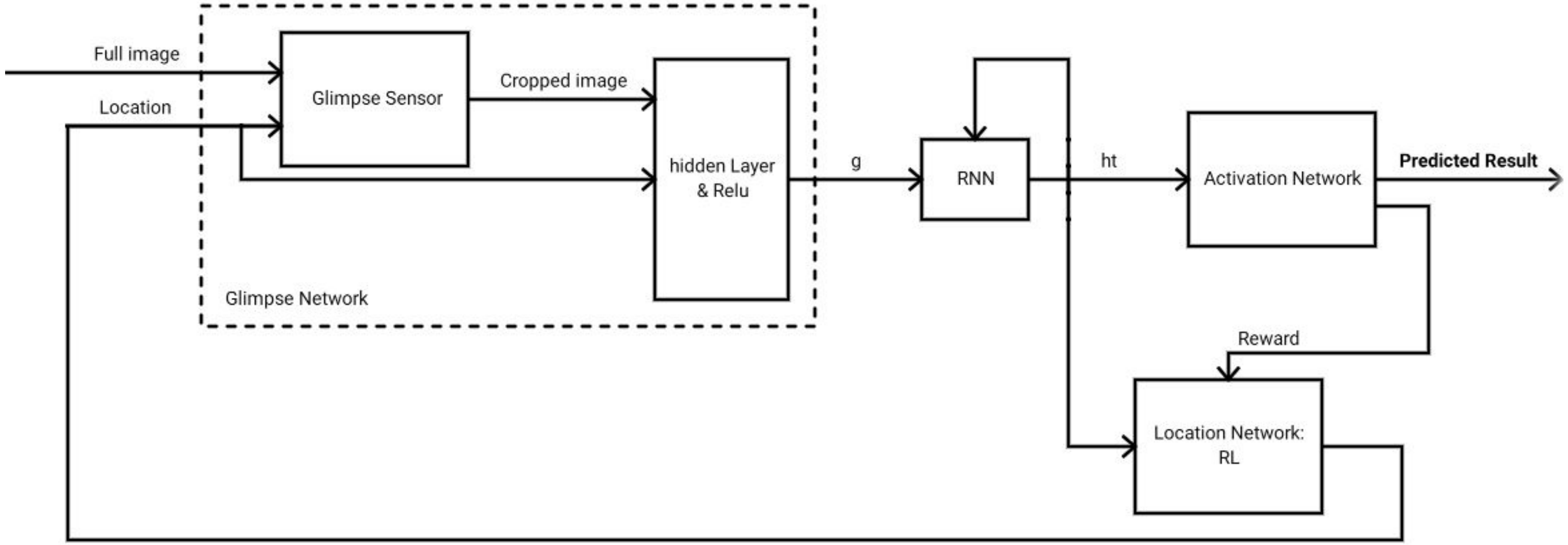


Figure 6: Neural Attention Model flow chart

3. Hyperparameters

3.1 Model-Based Hyper-parameters

3 Hyper-parameters:

- Number of glimpses
- Number of Scales
- Sensor Bandwidth

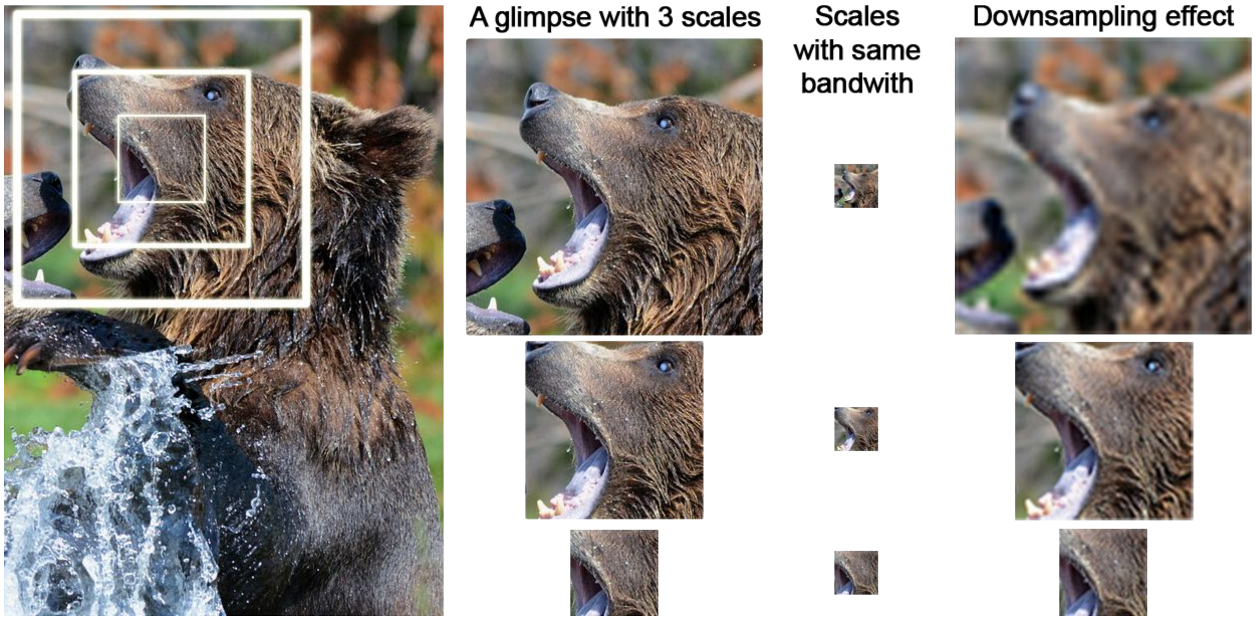


Figure 7: Example of 3 scales with a bandwidth of 44x44 px

3.2 Train-Based Hyper-parameters

Epoch

During one epoch, all the examples in the train set pass through the neural network.

Batch size

the neural network can not analyse all the data set at once. Therefore we divide the data and run mini batches instead.

4. Experiment

3.1 Augmented MNIST

MNIST:
- 70’000 grayscale images
- 60'000 for the training

Augmented MNIST:
- Additive noise
- Nb. are translated/rotated
- Image size: 100 x 100 px

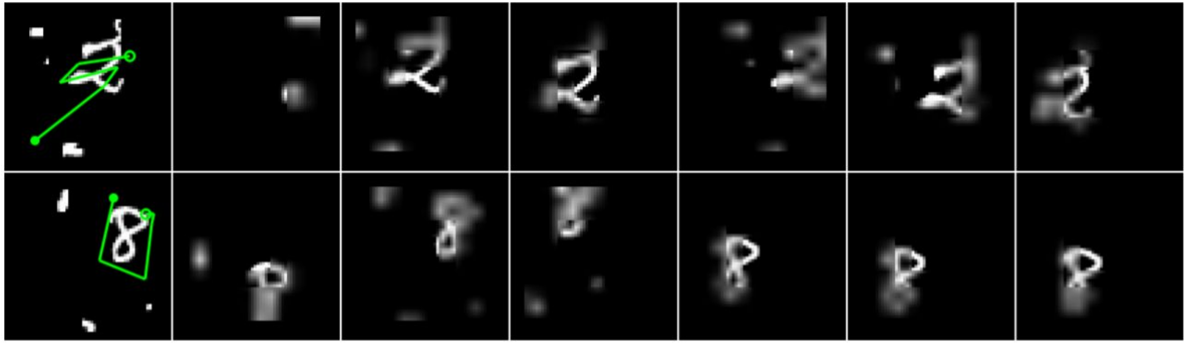


Figure 8: Examples of the learned policy on augmented MNIST images [8]

3.2 Experimental setup

- Environment :
  - Using Keras with TensorFlow backend → Visualisation using Tensor-board
  - GeForce GTX 1060 (compute capability: 6.1)
- Method:
  - linear search
- Result:
  https://github.com/Karim-53/Comparison-of-Neuronal-Attention-Models

5. Results

5.1 Number of glimpses

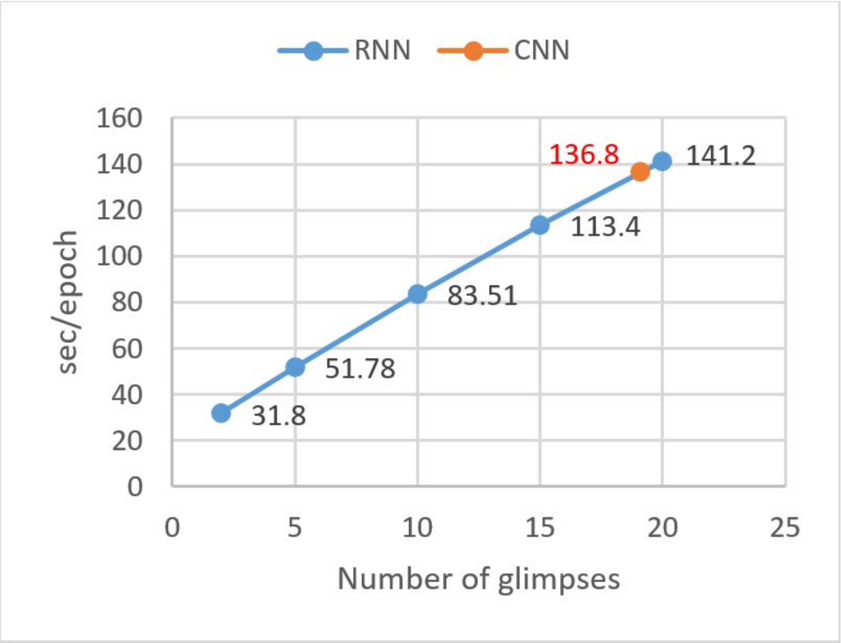


Figure 9: The training time on MNIST database using the baseline scenario and varying only the number of glimpses.

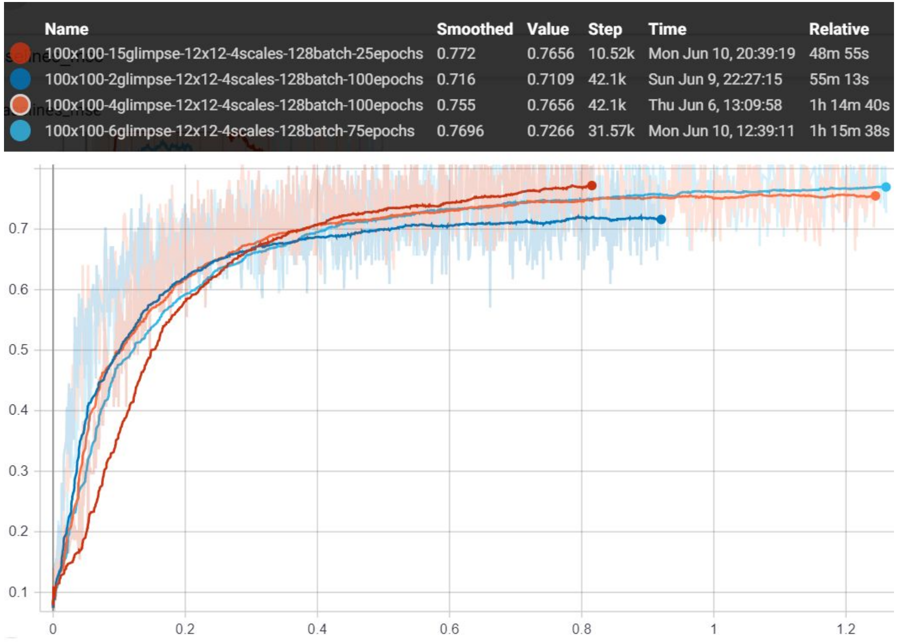


Figure 10: Evolution of the average accuracy through time while varying the number of glimpses.

5.2 Sensor Bandwidth

Sensor size = bandwidth² · number of scales

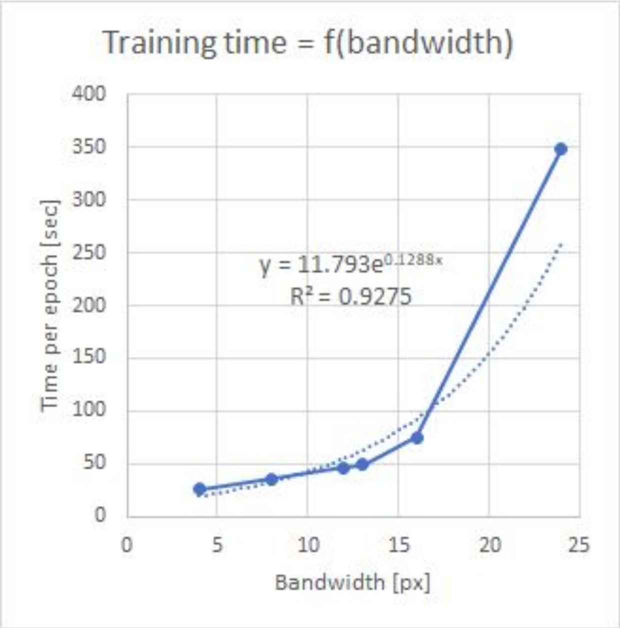
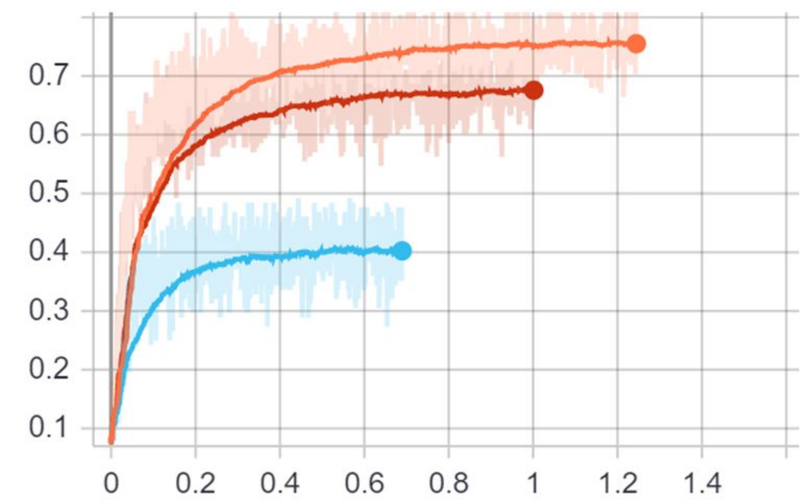


Figure 11: Evolution of the training time while varying the bandwidth



Name	Smoothed	Value	Step
100x100-4glimpse-12x12-4scales-128batch-100epochs	0.755	0.7656	42.1k
100x100-4glimpse-4x4-4scales-128batch-100epochs	0.4024	0.4219	42.1k
100x100-4glimpse-8x8-4scales-128batch-100epochs	0.6758	0.6875	42.1k

Figure 12: Evolution of the accuracy through time while varying the bandwidth

5.3 Sensor size

Sensor size = bandwidth² · number of scales

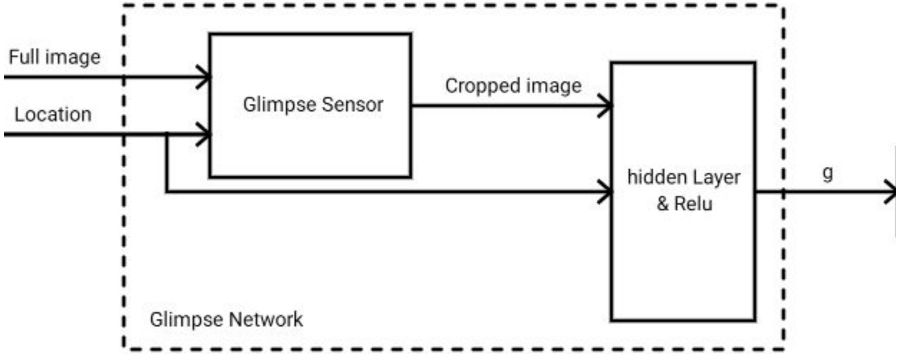


Figure 13: Neural Attention Model - partial flow chart

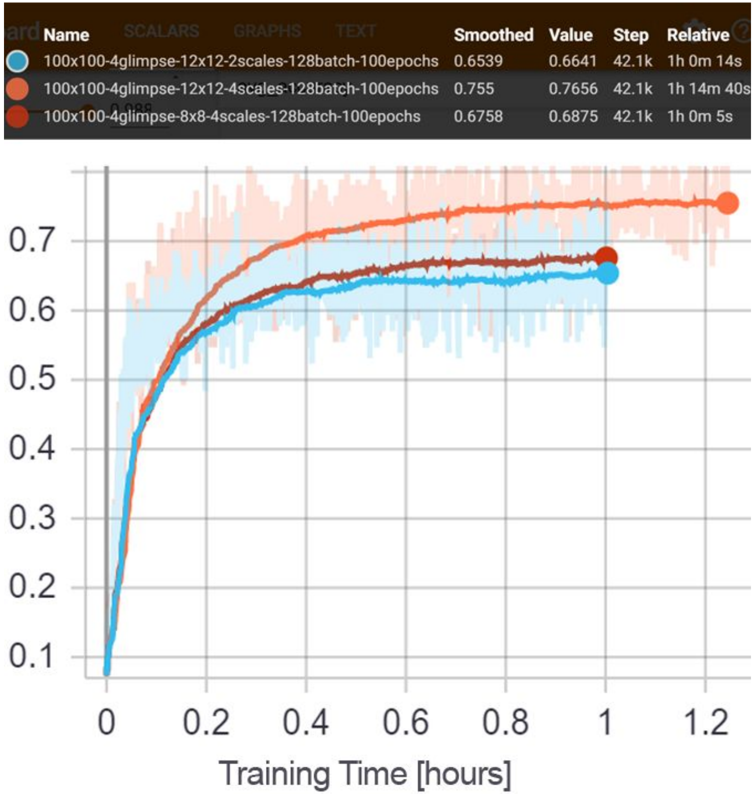


Figure 14: Comparing the effect of bandwidth and number of scales

5.4 Batch size

Optimal range:
64 - 256

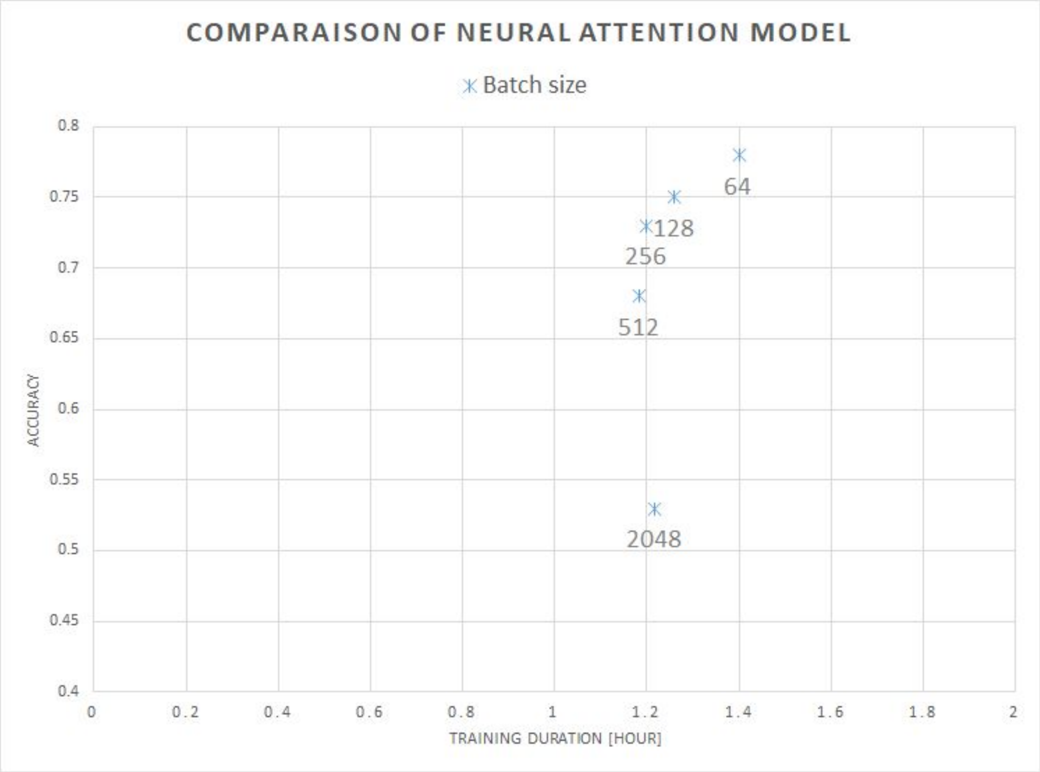


Figure 15: Effect of batch size on the accuracy and training duration

5.5 Number of epochs

# of epoch	Time (hour)	Accuracy
10	0.12	0.64
30	0.378	0.72
50	0.63	0.73
75	0.945	0.751
100	1.26	0.75

Table 1: Effect of the number of epoch on the accuracy and the training time

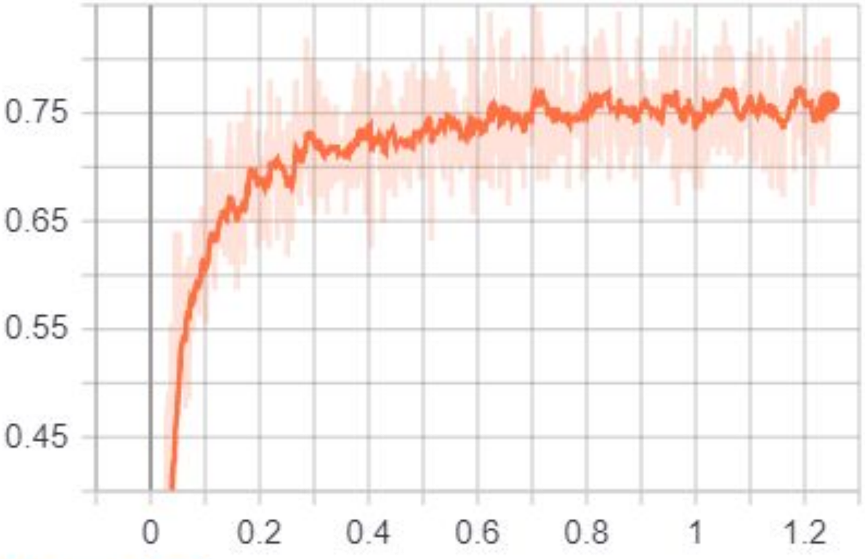
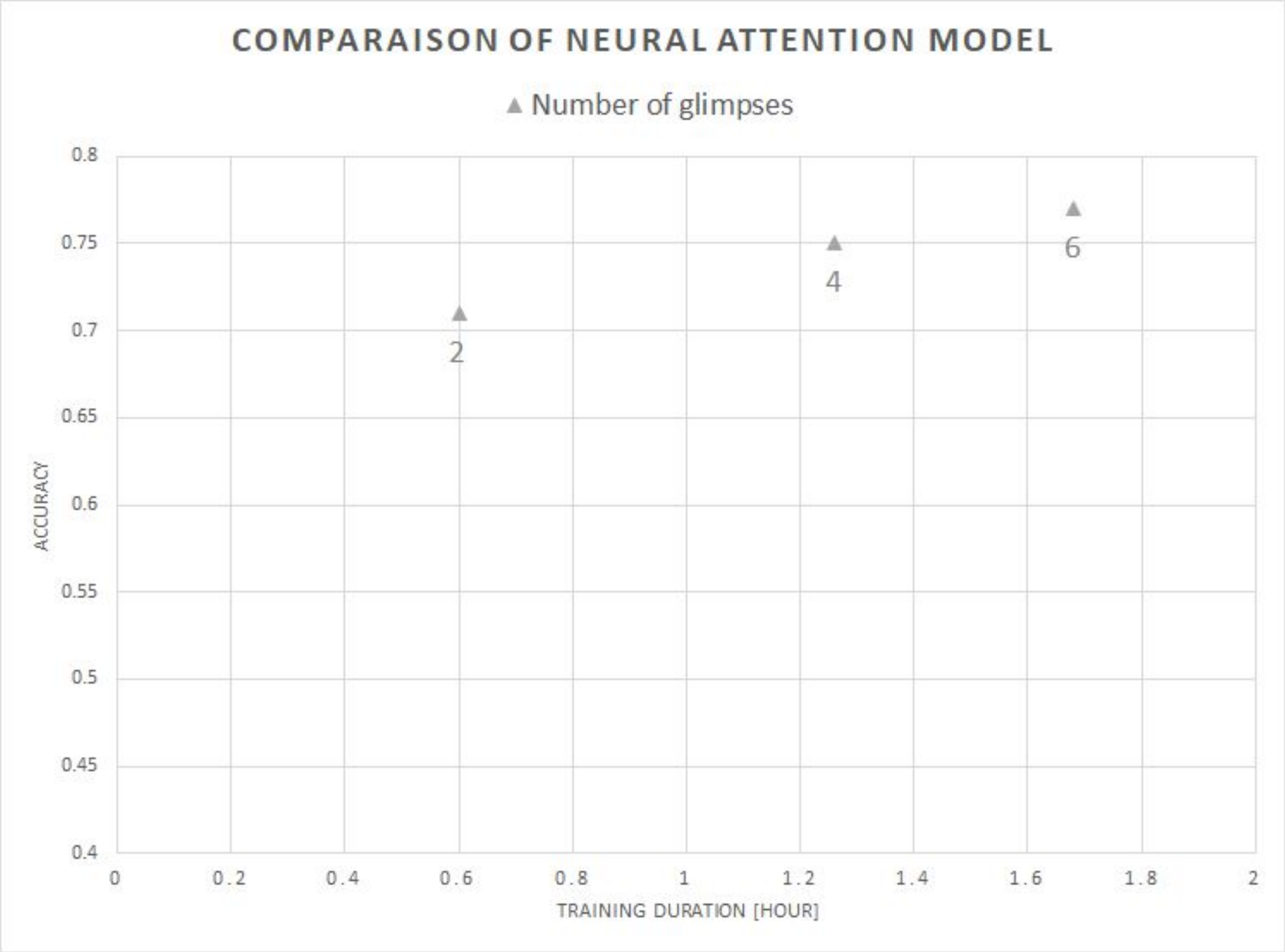
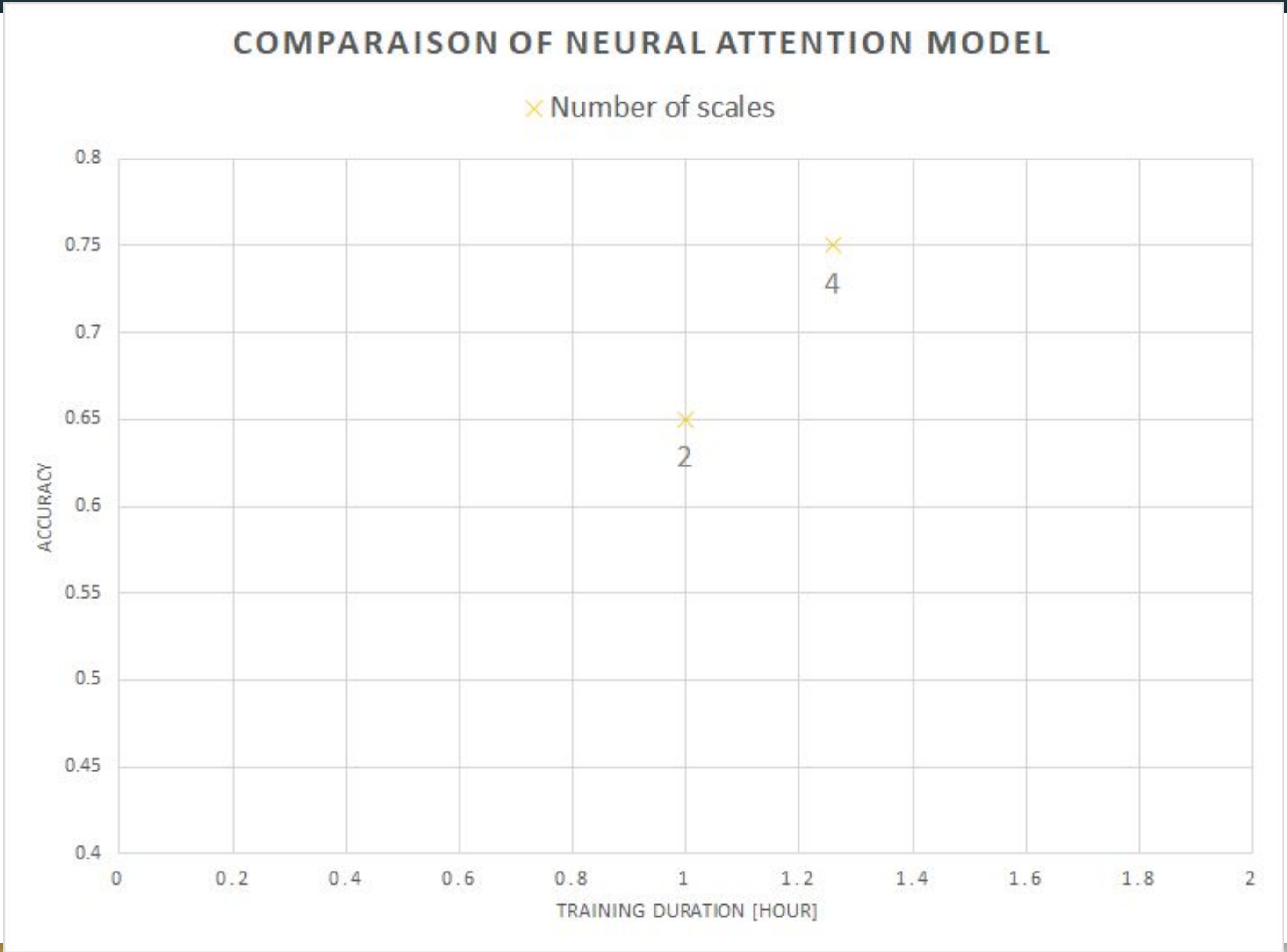
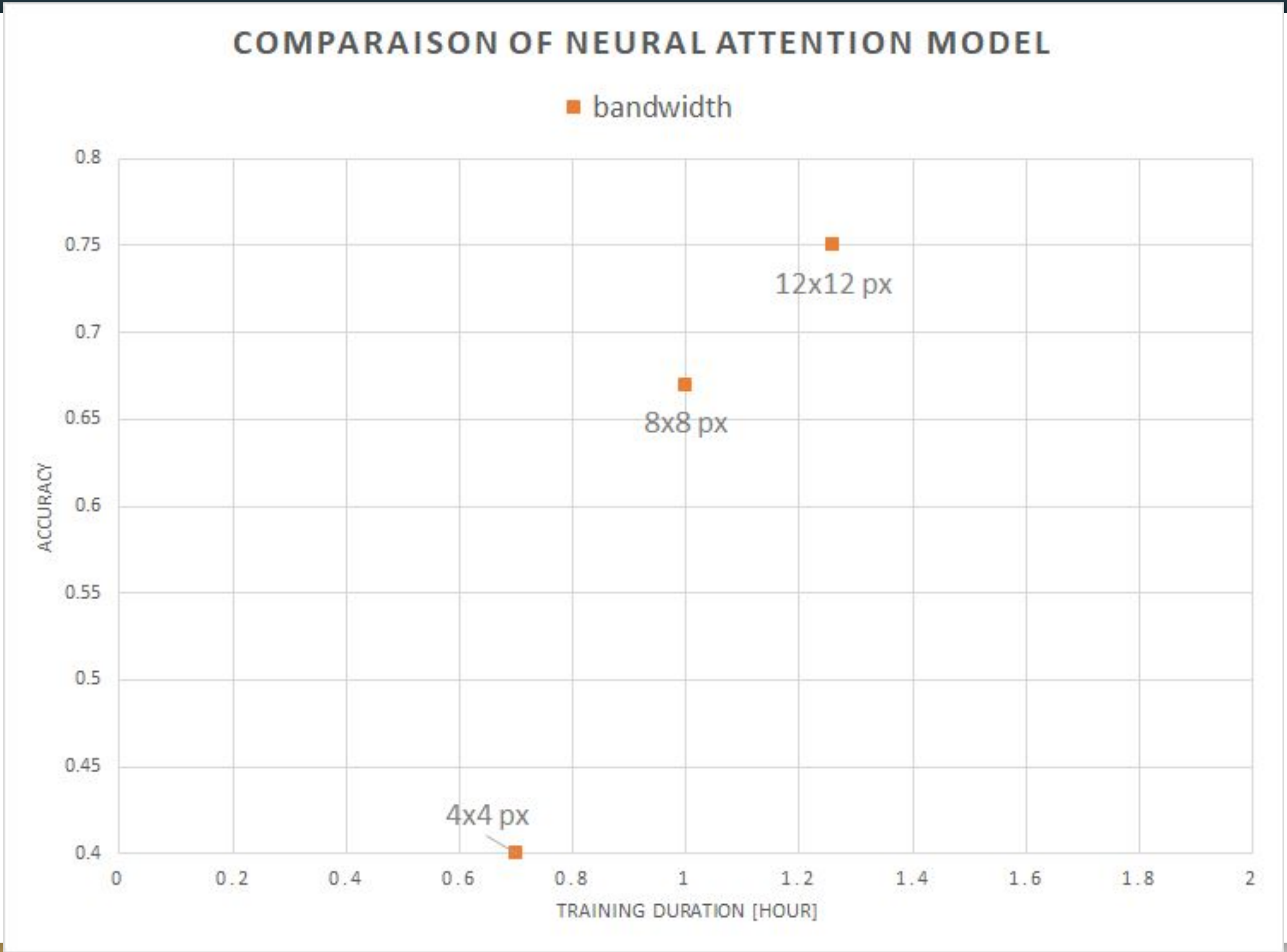


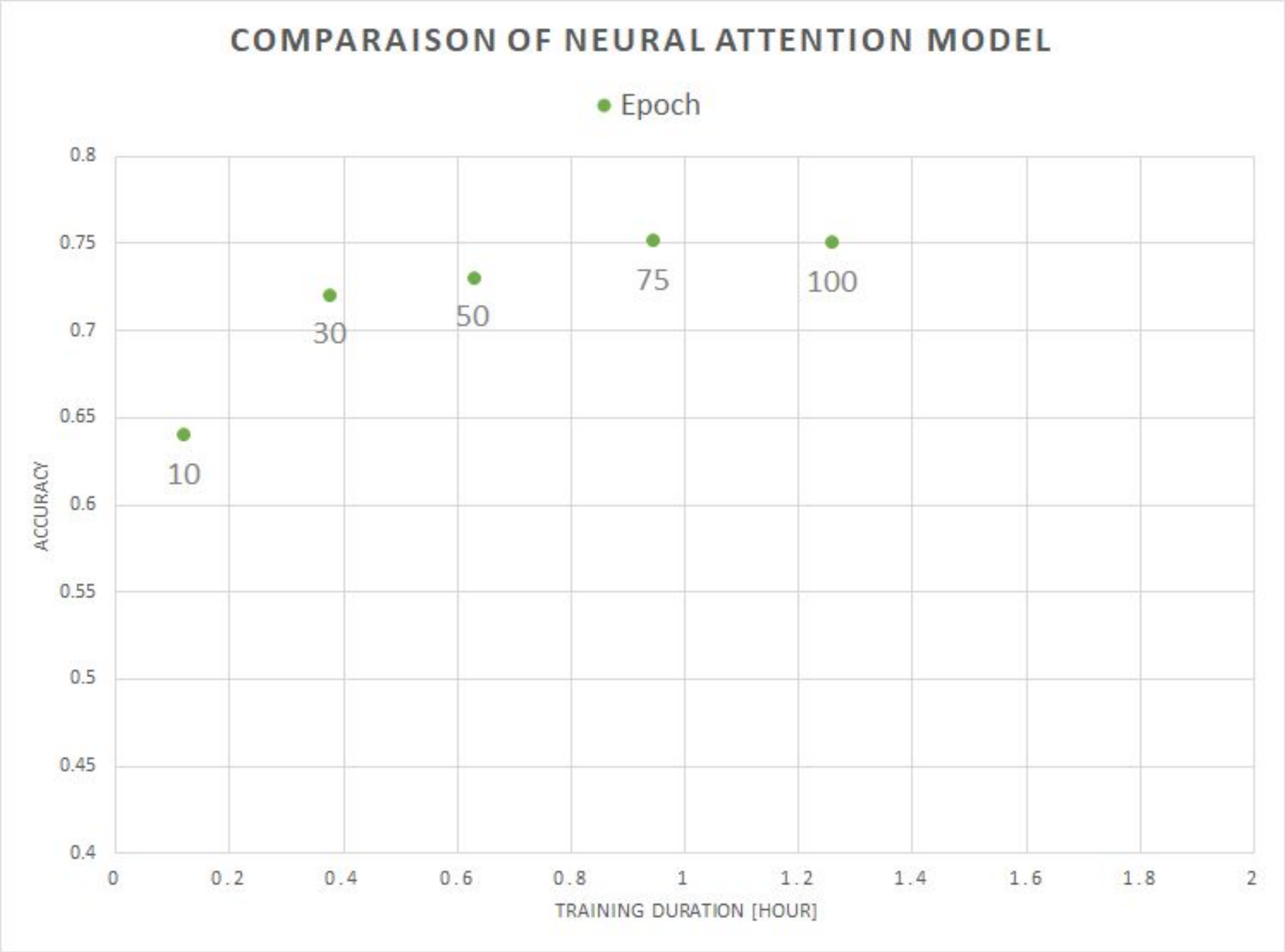
Figure 16: Evolution of the Accuracy

6. Conclusion

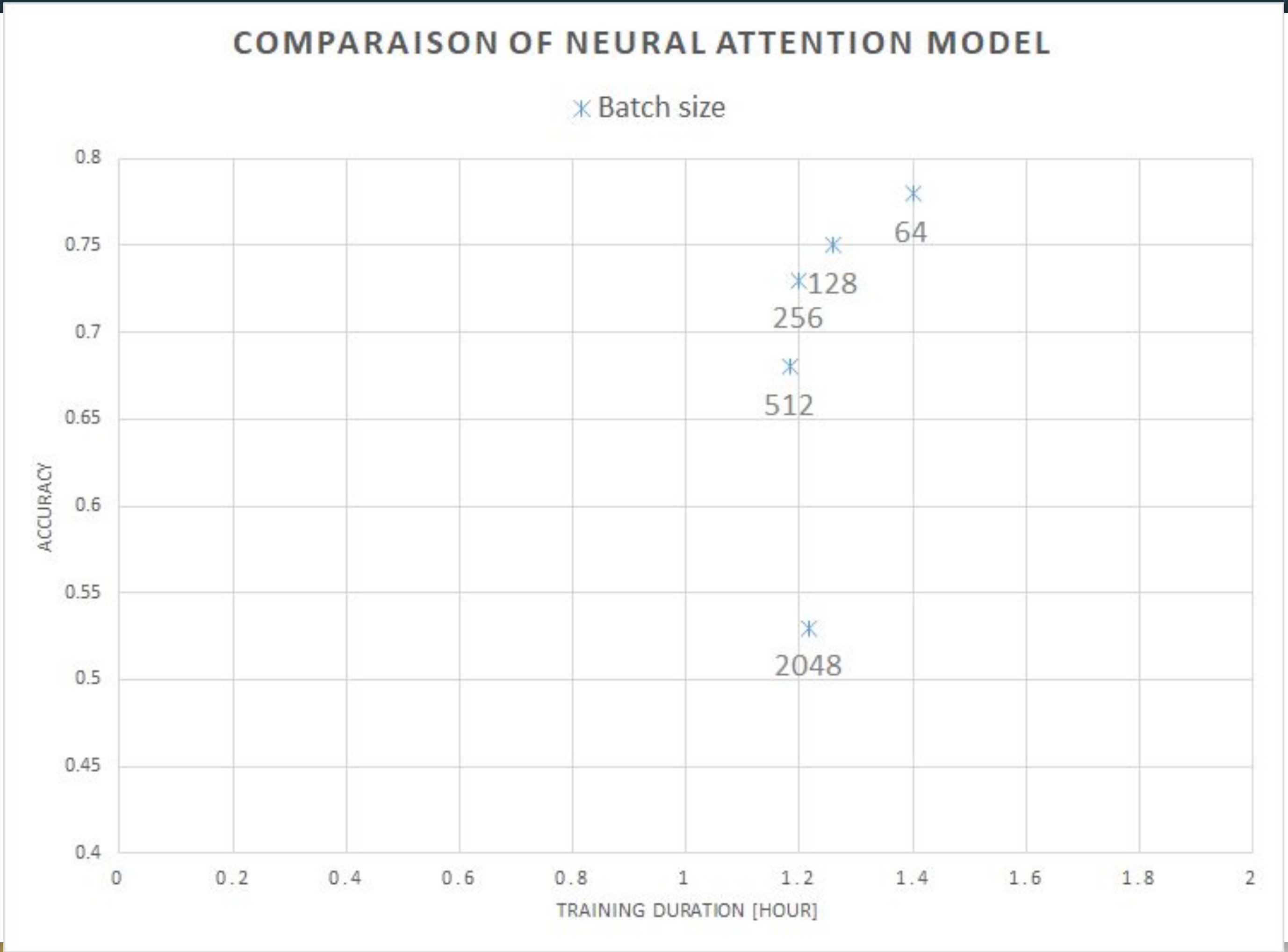

COMPARAISON OF NEURAL ATTENTION MODEL
Number of glimpses
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
ACCURACY
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
TRAINING DURATION [HOUR]
2
4
6


COMPARAISON OF NEURAL ATTENTION MODEL
Number of scales
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
ACCURACY
4
2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
TRAINING DURATION [HOUR]


COMPARAISON OF NEURAL ATTENTION MODEL
bandwidth
ACCURACY
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
12x12 px
8x8 px
4x4 px
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
TRAINING DURATION [HOUR]



COMPARAISON OF NEURAL ATTENTION MODEL
Epoch
ACCURACY
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
10
30
50
75
100
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
TRAINING DURATION [HOUR]


COMPARAISON OF NEURAL ATTENTION MODEL
Batch size
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
ACCURACY
64
128
256
512
2048
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
TRAINING DURATION [HOUR]

Thank you

References

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2016. TensorFlow: A System for Large-Scale Machine Learning. In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16). USENIX Association, Savannah, GA, 265–283. https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi

[2] Y.Y. BAYDILLI and U. ATILA. 2018. Understanding effects of hyper-parameters on learning: A comparative analysis. International Conference on Advanced Technologies, Computer Engineering and Science (2018). http://indexive.com/uploads/papers/icatces2018-175.pdf

[3] Mohamed Karim Belaid. 2019. Comparison of Neuronal Attention Models. https://github.com/Karim-53/Comparison-of-Neuronal-Attention-Models/.

[4] Ian Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, and Vinay Shet. 2014. Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks. In ICLR2014.

[5] Mary Hayhoe and Dana Ballard. 2005. Eye movements in natural behavior. Trends in Cognitive Sciences, Article 9 (2005), 188-194 pages. https://www.cs.utexas.edu/~dana/Hayhoe.pdf

[6] Koray Kavukcuoglu Jimmy Ba, Volodymyr Mnih. 2014. Multiple Object Recognition with Visual Attention. (2014). arXiv:1412.7755 https://arxiv.org/abs/1412.7755

[7] Yann LeCun and Corinna Cortes. 2010. MNIST handwritten digit database. http://yann.lecun.com/exdb/mnist/. (2010). http://yann.lecun.com/exdb/mnist/

[8] Volodymyr Mnih, Nicolas Heess, Alex Graves, and Koray Kavukcuoglu. 2014. Recurrent Models of Visual Attention. CoRR abs/1406.6247 (2014). arXiv:1406.6247 http://arxiv.org/abs/1406.6247

[9] Hiroyuki Shinoda Mary M Hayhoe Anurag Shrivastava. 2001. What controls attention in natural environments? Elsevier BV 41 (2001), 3535-3545. https://www.sciencedirect.com/science/article/pii/S0042698901001997#aep-bibliography-id21

[10] Keras Team. 2018. Keras example: MNIST CNN. https://github.com/kerasteam/keras/blob/master/examples/mnist_cnn.py.

[11] Tianyu Tristan. 2017. Visual Attention Model. https://github.com/tianyu-tristan/Visual-Attention-Model/tree/master/MNIST/ram.

[12] Tianyu Tristan. 2017. Visual Attention Model in Deep Learning. (2017). https://towardsdatascience.com/visual-attention-model-in-deeplearning-708813c2912c

[13] A. Yarbus. 1967. Eye Movements and Vision. Plenum Press (1967). https://www.sciencedirect.com/science/article/pii/S0042698901001997

[14] Adam Coates Alessandro Bissacco Bo Wu Andrew Y. Ng Yuval Netzer, Tao Wang.2011. Reading Digits in Natural Images with Unsupervised Feature Learning. NIPS Workshop on Deep Learning and Unsupervised Feature Learning