Perception of Digital vs. Analog Clocks by Deep Learning Networks

Kenneth Xu

Abstract

Many people find that while digital clocks give a more accurate reading, it takes less "mental energy" to perceive time by reading analog clocks. That is the basis as to why most people prefer analog clocks over digital clocks in daily life.

The objective of this paper is to find whether Artificial Neural Networks can replicate and confirm cognitive phenomena in clock reading. If the results are positive, the findings may increase our confidence in artificial neural networks modeling human visual perception in many visual psychological studies. For this purpose, the Average Absolute Neuron Output was proposed as a measurement for mental energy expended.

Two types of CNNs (Convolutional Neural Networks) were configured and trained with computer-generated clock images. The images embody digital clocks, analog clocks, as well as the patented Kid Klok.

For the first type of CNN, every layer is trained ground up to read the respective clock type. Results show that after sufficient training, both accuracy and "mental energies" converge to be around the same values.

For the second type of CNN, which more closely simulates the human vision system, the convolution layers are adapted from the VGG16 network. The results indicate that the digital clock readings are more accurate, while the perception of analog clocks and Kid Klok take less "mental energy" to process each image.

1. Introduction

The invention of digital clocks gives human beings the means of time reading with unparalleled accuracy [1]. However, people tend to prefer analog clocks over digital clocks in daily life as if perceiving time by reading digital clocks takes more "mental energy" in comparison to looking at the dials of analog clocks. The concept of spatial distancing is perhaps why the Analog Clock remains popular today. Spatial distancing allows humans to process time reading from a quantity rather than just an empty reading presented in the digital clock [17]. The Kid Klok [3] is proposed to be an improved version of the Analog Clock and is predicted to be easier to read. The difference in "mental energy" is proposed to be a likely result of the incorporation of exact roman numerals, separating the hour and minute dials, and the Chunking phenomenon [2] while maintaining spatial distancing.

It is proposed that perceiving time in the lens of spatial distancing is the more natural way of perceiving time because how much time passed is directly reflected in the magnitude of degrees the hands have changed [3]. This is one of the fundamental advantages of the long-lasting Analog Clock. However, while reading Analog Clocks, common errors can occur where readers mix the hour hand and the minute hand reading. An example of this is reading 2:15 as 3:10. The Kid Klok aims to alleviate this problem by separating the distinct hour and minute time dials. Another common problem is that people sometimes forget to multiply the minute reading by 5. An example of this is when someone reads 3:15 as 3:03. Dr. Massaro proposed that by incorporating the exact minute from 0-55, there would be less effort spent on the extra step of multiplying by 5 [14].

Another possible advantage the Kid Klok provides is through chunking. Chunking in cognitive psychology is a mental process whereby individual pieces within an information set are split up and then regrouped as a whole that's more meaningful [16]. In the specific case of analog clocks, all readings of different times are organized into a system of a dial and several hands. Compared with recognizing and memorizing the system of 10 digits, the dial/hand system is more compact and should be a more intuitive process to read.

In recent years, Artificial Neural Networks (ANNs) [4] have experienced numerous breakthroughs. It is well known that artificial neurons are derived from biological neurons. Figure 1 demonstrates the typical structures of a biological neuron and an artificial neuron. In the biological neuron structure, the input signals are weighted at the synapses and the dendrites pass the weighted input signals to the cell body. The cell body evaluates the sum of the weighted input signal of the ion gradient. If the sum passes a certain threshold, the cell activates and fires an output signal down the axon.

An artificial neuron is a mathematical model of the biological neuron. The numerical values representing the input signals are multiplied with different connection weights and later summed up together with a bias that corresponds to the aforementioned threshold. The sum is then fed into a nonlinear activation function. The output value of the activation function mimics the "firing" or "non-firing" of the biological neuron. ReLU (Rectified Linear Unit) activation function; $f(x) = \max(0, x)$ is one of the most popular activation functions used in ANN designs. Figure 2 shows a typical structure of artificial networks that mimic the interconnected biological neuron network. Each node in the ANN is an artificial neuron described in Figure 1 [13].

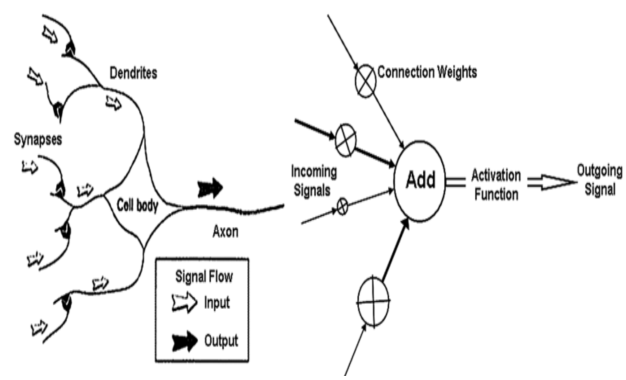

Figure 1. Structures of biological and artificial neurons

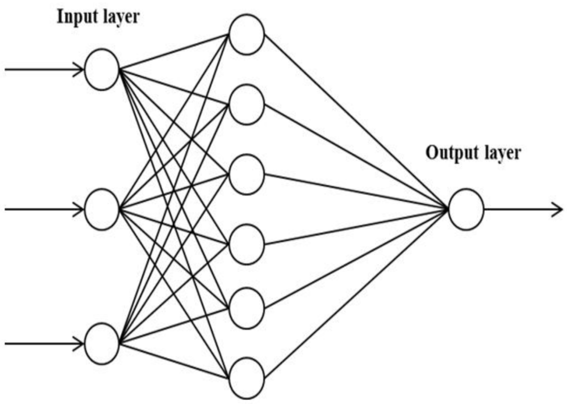

Figure 2. Structure of ANN with one hidden layer

In a biological neuron network, the more neurons are activated, the more work the neurons do and the more energy is expensed. Therefore, the total number of activated neurons, or the average possibility a neuron fires can be used to measure how much effort it takes the biological neuron network to finish a perception task. On the ANN side, the average possibility of neuron firing can be simulated by taking the average absolute output values of the artificial neurons. The higher the average, the more "mental energy" it requires to carry out the perception task. Therefore, we propose to use the average absolute neuron output (AANO) as the measurement of mental effort in this work.

One specific type of ANN, known as the Convolutional Neural Network (CNN), has proven to be particularly successful in processing visual data and other two-dimensional data. In tasks such as pattern recognition, CNN is approaching human-level performance [5]. A typical CNN (Figure 4) starts with multiple convolutional layers to extract image features such as corners, edges, specific-colored patterns, etc. The image features are passed into one or more full connection layers that will give these identified features meaning and generate final outputs.

Given the explosive developments of CNNs, the question arises whether we can observe the same discrepancies between digital and analog clocks readings, in terms of accuracy and "mental energy" consumption using CNNs. If this can be proven, not only do we have an indirect verification of the above psychological findings, but we may also have more confidence that a well-designed ANN is capable of simulating human brains adequately enough in many visual perception studies. Needless to say, testing with ANN is by far more efficient and less risky as compared to testing with living humans and human brains.

Using CNN to read clock images is not a new concept. S. Verma [6] and F. Duvallet [7] have designed CNNs to read analog clocks in the past. Yu et al. have researched using CNN to read digital clocks in video clips [8]. Their works have all proved that CNNs are capable of reading clocks with human accuracy. However, no one has ever used CNNs to study the accuracy and mental effort of different clock types.

In this research, two types of CNN were used to study the digital and analog reading differences. While both were trained to read the three clock types (analog clock, digital clock, Kid Klok), the second type CNN was much more representative of a human visual perception system. This is because the second type CNN includes layers that were trained for a variety of other visual tasks, much like a human being. In comparison, the first type CNN is not as representative because it was trained only for the visual task of Clock Reading.

It should be noted that although the inference by the CNNs could be a good reflection of perceiving time by taking a glance at a clock, the CNN training itself is not a simulation of the learning process of clock reading by children. The training of the CNNs is based on the Backpropagation algorithms [15], which are solely designed to adjust CNN parameters to minimize Loss functions. On the other hand, children's time learning not only requires practice, but it also involves many more brain activities such as reasoning, association, and mental calculation. This process is way more complex and is beyond the scope of the simple CNNs used in this research.

2. Method

For each type of clock, Scilab scripts are used to generate 720 clock images. These images cover every reading possible from 00:00 to 11:59. Therefore, 2160 total images are generated (720 per clock type). Figure 3 displays a sample image of each clock type. To balance training speed and clock reading accuracy, the final resolution of the pictures is set to be 160x160. For each group of 720 images, 600 images are randomly picked for training. The remaining 120 images are saved for testing.

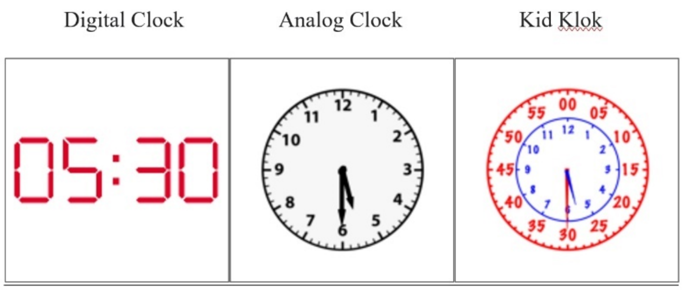

Figure 3. Sample 160x160 training images of each clock type (CNN Type 1)

For the first type of CNN, the design of S. Verma is being adopted. The only modification is made to match the receiving size of 160x160x3 instead of the original 100x100 grayscale single channel input. The overall framework of the network is visualized in Figure 4.

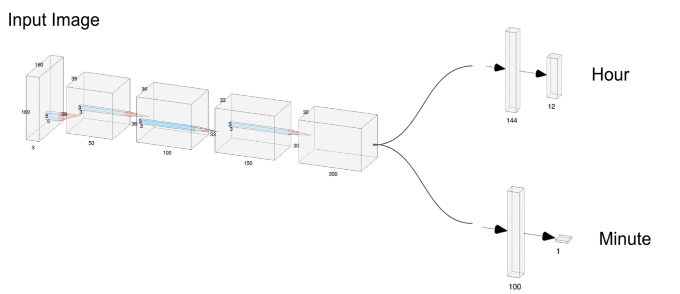

Figure 4. Architecture of CNN Type 1

A typical CNN network [5] has two types of layers, convolution layers, and full connection layers. The convolution layers are responsible for extracting image features, such as edges, corners, color patterns, etc. The full connections take the image features as the input and "perceive" meaning out of them.

In CNN Type 1 (Figure 4), there are a total of four convolutional layers. The activation functions of the convolutional layers are ReLU and every convolutional layer is followed by 2x2 Maxpooling except for the last layer. The last convolutional layer is followed by a dropout layer to control overfitting. The network forks into two branches split between the Hour branch and the Minute Branch after the dropout layer. The Hour Branch has two full connection layers. The first full connection layer has 144 neurons. The second full connection layer includes 12 neurons that are also the output nodes. As stated in Verma's work, we treat the Hour Reading as a classification task. Minute Reading, however, is a regression task. The Minute Branch also has two full connection layers. The first full connection layer includes 100 neurons, and the second layer includes only one neuron that gives the minute value output. Therefore, 4 full connection layers include a total of 257 neurons to "perceive" time in CNN Type 1.

For each type of clock reading, the network is trained ground up. The parameters of every layer are updated during training. The selected learning rate is 0.001 and the batch size is set at 60. In terms of the loss function, a combination CrossEntropyLoss for Hour classification and MSELoss for Minute regression is utilized. For every clock type, the network is trained for 200 epochs before testing.

The average absolute neuron output (AANO) of all 257 neurons in full connection layers which do the "perception" job is employed as the measurement for expendable "mental energy". As explained in the Introduction section, a high AANO means that more neurons are activated during the perception process. The more neurons activated; the more energy consumed. In other words, AANO is positively correlated to "mental energy" consumption.

Designing and training a CNN solely for clock reading does not simulate human visual perception accurately enough. The human vision system is used to perform thousands of different perception jobs, and reading clocks are just one of them. Therefore, the convolutional layers within the CNN should be at least geared towards the general inclusive vision as opposed to only clock reading. With this in mind, the second type of CNN is introduced, and Transform Learning is utilized as opposed to ground-up learning.

In the second type of CNN (Figure 6), the first type's convolution layers are replaced with the convolution layers in the VGG16 network [10]. In these layers, the parameters are fixed throughout clock reading training. The full connection layers remain consistent with the first type of CNN. We still use 4 full connection layers that include 257 neurons to "perceive" time and the AANO computation is consistent. The parameters of all 4 full connection layers are updated from the ground up during the training for each clock type.

The VGG16 network was originally designed for the task of image classification of up to 1000 classes. Its parameters have been optimized by millions of pictures from ImageNet [11]. VGG16's convolutional layers have been popular for Transform Learning tasks over the years, and its convolutional layers are suitable for the clock reading network to mimic human vision systems tuned for general perception jobs.

The architecture of the second type of CNN is demonstrated in Figure 5. It is noticeable that VGG16 takes 224x224 RGB images as input, rather than 160x160 RGB images. To mend the resolution difference, every 160x160 image is padded with 32 lines on each side in order to generate the 224x224 images. Sample images of the second type of CNN are illustrated in Figure 6.

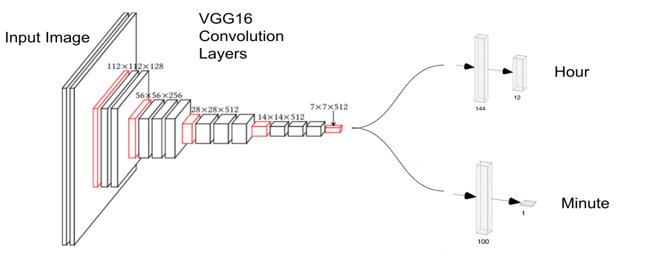

Figure 5. CNN Type 2 Architecture.

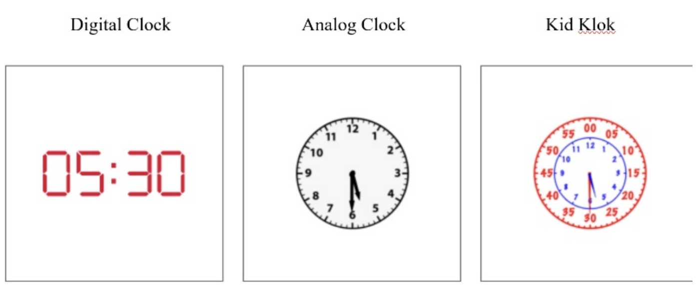

Figure 6. Sample 224x224 training images of each clock type (CNN Type 2).

The construction of CNNs and their training are performed in the Pytorch framework [12]. All the Pytorch code along with the Scilab scripts for image generation can be downloaded from https://github.com/Kenneth-Xu11566/Clock_Reading_By_CNN

3. Data/Results

3.1 CNN Type 1

Since Hour Reading is a task of classification, accuracy is reflected as the rate of being correct in terms of percentage. The Minute Reading is a problem of regression, so the accuracy is measured in standard deviation from the ground truth.

Figure 7, Table 1, Figure 8, and Table 2 are results of the first type of CNN.

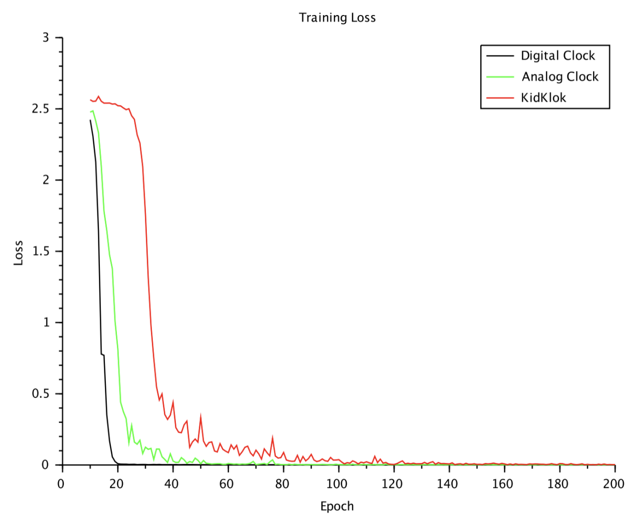

Figure 7. Training Loss vs. Epoch (CNN Type 1).

	Digital Clock	Analog Clock	Kid Klok
Hour Reading Accuracy (Correct Rate)	100%	99%	98.3%
Minute Reading Accuracy (Standard Deviation)	0.513min	1.008min	1.009min

Table 1. Clock Reading Accuracy on test datasets.

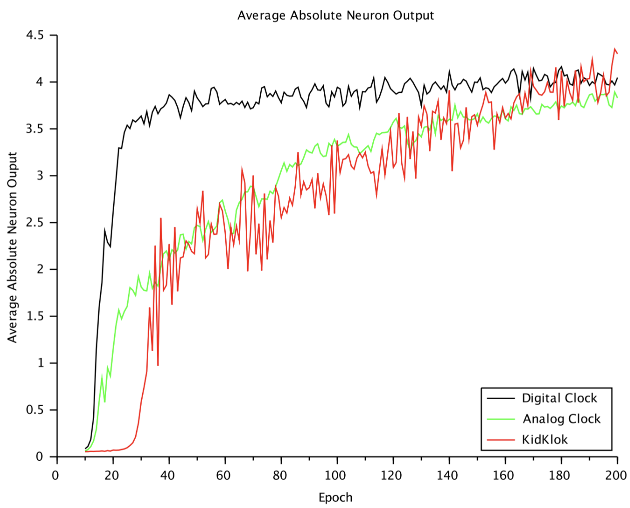

Figure 8. AANO vs. Epoch (CNN Type 1).

	Digital Clock	Analog Clock	Kid Klok
Average AANO of final 10 epochs	4.018	3.831	4.054

Table 2. Average AANO of the final 10 epochs.

Figure 7 illustrates the change of Losses of all three types of clocks reading training relative to the training epoch count. The plot indicates that despite the Digital Clock Loss dropping sooner, all three clocks seem to approach a very low value by the end of the 200 epochs. This means that after training, all three types of clock readings are fairly accurate.

Table 1 depicts the accuracy of all three types after testing with the test datasets. One can argue that the digital clock has a slight advantage. However, the advantage is yet unclear because all three types of clocks reading by the trained CNN are very accurate and statistically indistinguishable.

Figure 8 shows the AANO change of the three types of clock readings over 200 training epochs. It can be concluded that by the end of the training, the AANOs of all types of clock reading are similar.

Table 2 demonstrates the average AANO values of the final 10 epochs. Evidently, the values are similar.

3.2 CNN Type 2

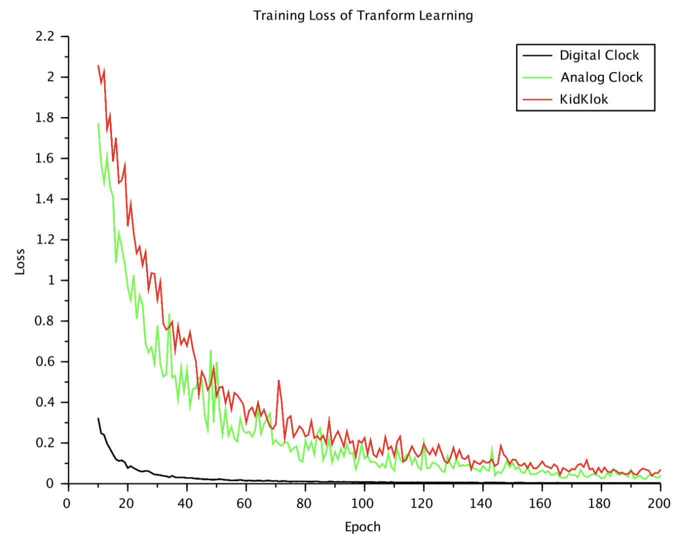

Figure 9. Training Loss vs. Epoch (CNN Type 2).

	Digital Clock	Analog Clock	Kid Klok
Hour Reading Accuracy (Correct Rate)	100%	92.5%	95.8%
Minute Reading Accuracy (Standard Deviation)	0.743min	4.295min	4.100min

Table 3. Clock Reading Accuracy on test datasets.

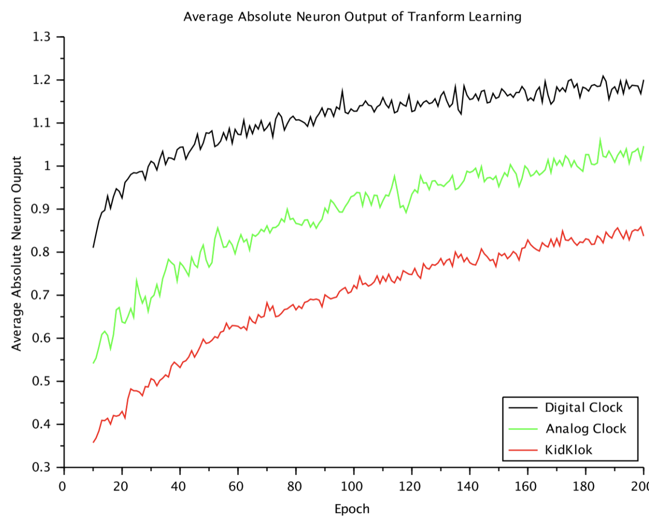

Figure 10. AANO vs. Epoch (CNN Type 2).

	Digital Clock	Analog Clock	Kid Klok
Average AANO of final 10 epochs	1.184	1.030	0.846

Table 4. Average AANO of the final 10 epochs.

Figure 9, Table 3, Figure 10, and Table 4 are the results of the second type of CNN.

Figure 9 depicts the change of Loss of all three types of clocks reading training relative to the training epoch count. The plot shows that while training progresses, all Losses are dropping. However, the Loss of digital clock training drops much sooner than the other two clock types. This suggests that after training, the digital clock reading is significantly more accurate than that of the other two clock types.

In comparison with Figure 7, the Losses displayed in Figure 9 drop at a slower rate. This is reasonable because in the second type of CNN, the parameters of convolutional layers' parameters are fixed, while the first type of CNN's convolutional layer parameters is also being updated for clock reading. Therefore, the accuracy progress is slower in the second type of CNN.

Table 3 describes the accuracy measurement of all three clock types after testing with the test datasets. It is clear that the accuracy of digital clock readings is superior to the other two types of clock readings. This is especially notable on the Minute Reading, where the digital clock is over 5 times more accurate than both the Kid Klok and the Analog Clock.

Figure 10 illustrates the AANO change of the three types of clock readings in conjunction with the training epoch count. We can see that the AANO of Kid Klok is always lower than that of the Analog Clock. Interestingly, the AANO of the Analog Clock reading is always lower than that of the Digital Clock. This implies that if the clock reading is not required to be exactly accurate, the mental energy for Analog Clock reading is always smaller than that of the Digital clock, while the effort for Kid Klok reading is the lowest.

Table 4 depicts the average AANO of the last 10 epochs. The numbers indicate the same analysis.

4. Conclusion/Discussion

The results of the first type of CNN do not clearly expose the psychovisual phenomena mentioned in the Introduction. The second type of CNN, however, indeed confirms those observations. The readings of digital clocks are truly significantly more accurate but present the highest required effort, while the readings of the Kid Klok take the least effort/mental energy but with lower accuracy. The analog clock was found to be the median in terms of effort, and on par with the Kid Klok in terms of accuracy. Therefore, it can be concluded that well-crafted CNNs are capable of simulating human visual perception systems in clock reading cases. The benefits are significant, with artificial neural networks being more cost-effective and more efficient to run studies with.

On the other hand, it cannot yet be firmly concluded that human subjects in visual psychological studies can be completely replaced by ANN as only one type of case study is resolved in this paper. No firm conclusion can be drawn until we study a wider variety of cases. It is also worthy to remember that much of the results depend on how the networks are designed and trained, so there is a certain inescapable degree of variability.

As discussed in the Introduction, in this research, we are only simulating time perceiving, not simulating the learning process of clock reading by children, which is a much bigger topic. That being said, designing an ANN to simulate the process of clock reading learning could be a good direction for future study.

One may also notice how all the training images are clean, well centered, and strictly upright. In reality, the inputs to the human vision system are rarely this perfect. What we see tends to be tilted, of different sizes, and under the influence of all different lighting conditions. As a future study, a lot more training images that are closer to what we see in real life can be included.

5. Acknowledgments

I thank Dr. Chen for their supervision and support. His technical knowledge and expertise were invaluable to the inception and continued development of this work. I also must thank Dr. Massaro for not only letting me use Kid Klok images to train and test CNNs, but for reviewing the draft and providing insightful advice. The training and testing images of digital clocks and regular clocks were derived from the online pictures on

https://timer.onlineclock.net/andhttps://www.pngitem.com/middle/hmoRoiJ_1-05-clock-clipart-clipart-stock-clipart-clock/

6. References

[1] Chintala, Soumith (1 September 2016). "PyTorch Alpha-1 release". retrieved on August 24, 2021, from https://github.com/pytorch/pytorch/releases/tag/v0.1.1

[2] "CogniFit." Spatial Perception- Cognitive Ability, 18 Feb. 2016, www.cognifit.com/science/cognitive-skills/spatial-perception.

[3] D. Massaro, "Child's Easy-to-read Timepiece" U.S. Patent 4,885,732, issued December 5, 1989

[4] Deng, Jia; Dong, Wei; Socher, Richard; Li, Li-Jia; Li, Kai; Fei-Fei, Li (2009), "ImageNet: A Large-Scale Hierarchical Image Database" (PDF), 2009 conference on Computer Vision and Pattern Recognition

[5] Felix Duvallet (2016), Deep time (using Tensorflow to read clocks), retrieved on August 24, 2021 from https://felixduvallet.github.io/blog/deep.time/

[6] Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron (2016). "6.5 Back-Propagation and Other Differentiation Algorithms". Deep Learning. MIT Press. pp. 200–220. ISBN 9780262035613.

[7] Gobet, Fernand. "Chunking Mechanisms in Human Learning." 2001.

[8] J. Brownlee (2017), A Gentle Introduction to Transfer Learning for Deep Learning, retrieved on August 24, 2021 from https://machinelearningmastery.com/transfer-learning-for-deep-learning/

[9] K. Simonyan, A. Zisserman (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition arXiv:1409.1556v6 [cs.CV]

[10] Massaro, D., & Rothman, D. (2002). Time to Learn About Time. Fuzzy Books. Schwartz, R. (2008). The Birth of the Meter. Math Horizons, 16(1), 14-31. Retrieved July 23, 2021, from http://www.jstor.org/stable/25678768

[11] Massaro, Dominic. "Overcoming Difficulties in Telling Time." Psyentific Mind, 1991, retrieved on August 24, 2021, from http://psyentificmind.com/kid-klok-information/overcoming-difficulties-in-telling-time

[12] Memory Loss and Gains of Older Adults, Retrieved August 24, 2021, from https://s3-eu-west-1.amazonaws.com/www-documents/Memory+in+older+age_v2.pdf

[13] Shiva Verma (2020), Reading Clocks Using Neural Nets, retrieved on August 24, 2021, from https://towardsdatascience.com/training-neural-net-to-read-clock-time-9473175171e3

[14] Swain, Debadatta; Ali, MM; Weller, Robert, "Estimation of mixed-layer depth from surface parameters", 2006 Journal of Marine Research, 64(5)

[15] X. Yu et al (2015), Reading Digital Video Clocks, International Journal of Pattern Recognition and Arti¯cial Intelligence Vol. 29, No. 4 (2015)

[16] Yann LeCun (2016). Slides on Deep Learning, retrieved on August 24, 2021, from https://indico.cern.ch/event/510372/attachments/1245509/1840815/lecun-20160324-cern.pdf

[17] Zell, Andreas (2003). "Chapter 5.2". Simulation neuronaler Netze[Simulation of Neural Networks] (in German) (1st ed.). Addison-Wesley. ISBN 978-3-89319-554-1. OCLC 249017987.