POZNAN UNIVERSITY OF TECHNOLOGY ACADEMIC JOURNALS
No 104 Electrical Engineering 2020

DOI 10.21008/j.1897-0737.2020.104.0001

Sebastian KULA*

COOLING FAN CONTROLLED BY EMBEDDED VISION SYSTEM

The HMI (human machine interaction) systems are widely used to control machines and variety of devices. Currently the HMI solutions, based on touch screens are almost commonly used in many domains, however the number of devices, which interaction with the user is based on speech recognition or user gesture recognition increases systematically. The paper focuses on the electromechanical system, which applies gestures and handwritten digits to control the speed of the DC cooling fan. The system crucial elements are the AVR microcontroller and the developer board, equipped with the embedded supercomputer NVIDIA Jetson TX1. To create the software part of the system artificial intelligence algorithms and deep neural networks were applied. The paper describes the complete routine of data preprocessing, deep neural network training and testing with the use of the GPU Tesla K20 and with the use of the DIGITS (Deep Learning GPU Training System), deployment of the trained model on Jetson TX1 board and the system execution. The system enables to control the fan through the two gestures ("stone", "paper") or through four handwritten digits.

KEYWORDS: computer vision, deep neural networks, electromechanical systems, human computer interaction.

1. INTRODUCTION

Artificial intelligence methods are present in engineering since many decades and neural networks, which are essential for the AI (Artificial Intelligence) were designed in form of mathematical description in the first half of twentieth century [1]. Despite the concept of neural networks is relatively old, we observe dynamical increase of the AI applications in many engineering domains, which includes also electrical engineering, mechatronics and robotics. The development of AI is feasible due to appearance of cost effective and computational efficient parallel architectures, especially devices based on GPU (Graphical Processing Unit) cards. The GPUs are suitable to train neural networks, including DNN (Deep Neural Networks) as training routines can be transform into matrix operations. The matrix operations have great potential of parallelization, it results that execution of the operations on GPUs are much faster than on CPUs.

* Kazimierz Wielki University, Bydgoszcz

The AI is also applied in computer vision, the research domain strongly present in mechatronics and robotics. The recent development of computer vision is also strongly connected with the application of GPU cards, which are capable to effectively train neural networks with massive datasets. The dominance of GPUs in the CV (Computer Vision) and AI domain was proven in 2015, when the model, created for images classification recognized and classified images faster and better than human beings, those results were commonly announced as a breakthrough and the "Big Bang" in AI and CV [2]. For training the revolutionary successful images classification model authors used Nvidia K20 and K40 GPUs, and the dataset contained about 1.2 million training images, 50,000 validation images, and 100,000 test images [2].

Image classification systems enable to control robots with the use of HMI (Human Machine Interface) and through video camera. The goal of the paper is to present vision based electromechanical system, which control the speed of the cooling fan. The embedded vision systems are currently getting more and more importance in automotive applications and in home automation, this trend is proven in [3] and in other articles. In [3] the control system of home appliances is described, the system uses camera and routines of the gesture recognition. In [4] authors presented the real-time, vision based system for human-machine interface applied in cars, the model uses gestures.

The article presents complete routine of the gesture controlled system development. The routine is based on the DNN and the machine learning. The description of the routine starts with selection of data for training and validation, then hyperparameters settings of the DNN are presented. In the next step the explanation of the training process and training results, obtained with the use of GPGPU and DIGITS software are presented. Then the deployment of the trained DNN model on embedded Jetson TX1 board and the applied code, which was written in CUDA C and C language are depicted. Finally applied hardware and internal dataflow of the system are presented. The system is universal, it is proven by the design of two versions of the system. The first version is based on gestures and the second one is based on visual recognition of handwritten digits.
It is expected further development and quantity increase of the systems, which are based on gestures and image classification due to the evident observation, that image, seeing, gesture are simple and natural methods, for human beings to communicate, transmit and receive information.

The paper novelty is an application of the GPU embedded board and DIGITS software to design fan cooling system, which exploits significantly machine learning techniques. The paper is a written version and essential extension of the poster presented at the NVIDIA's GPU Technology Conference (GTC) Europe in 2018.

2. DATASET AND DNN APPLIED FOR IMAGE CLASSIFICATION

The advantage of the DNN training is ability to obtain correct models based on raw data, there is no necessity to prepare the data in the complex manner, like for example there is neglectable to convert RGB data into grey scale data.
To create the model, which is based on the DNN training two sources of datasets were used: the author's dataset, which were acquired with the use of Jetson TX1 development board camera and dataset available at the website [5, 6]. Both datasets contain two different categories of gesture images, the "stone" category, the example is shown on Fig. 1, the "stone" gesture in the prototype is responsible for stopping the fan and the "paper" category, the example is shown on Fig. 2, the "paper" gesture is responsible for accelerating the fan to the maximum speed. Images are colorful (24 bits per pixel), they have resolution 640x480 DPI (dots per inch), the files size vary from 400 kB to 466 kB. The website dataset was created with the use of four persons, which were photographed in the office during the day and natural lightening [5, 6]. The total number of images for training and validation was 312. The same amount of images was used for each classification category (for "stone" category 156 images and "paper" also 156 images). For validation 78 images (25% of dataset) and for training 234 images were used.

Fig. 1. The gesture of the "stone", which is used to stop the fan

Fig. 2. The gesture of the "paper", which is used to run the fan with the maximum speed

Except the dataset with images to create the model, the architecture of neural network is needed. The architecture enables to determine the weights, that is to train the neural network. Currently, the best results in images classification have been achieved with the use of CNN (Convolution Neural Networks) architecture. In the paper the CNN were also used, in concrete the AlexNet [7] and GoogLeNet [8] networks, both networks gave creditable results.

The AlexNet was designed in 2012 and originally the code was written in CUDA C and implemented for GPGPU device [7]. The GoogLeNet was created in 2014 and has much less parameters than AlexNet [8].

3. MODEL TRAINING IN DIGITS

Due to successful realization of the CNN on GPGPU devices it is currently norm to support CNN and DNN training with the use of the GPGPU. There are many tools to handle the training routines on GPGPU, one of such tool is DIGITS (NVIDIA Deep Learning GPU Training System). The DIGITS is the interactive tool for datasets, models management, neural network training, the tool enables to observe accuracy and training loss during the DNN training, it enables model validation through loading the data from internet or local desktop [9]. The DIGITS is a tool for images classification, image segmentation and for objects detection on images. In the article the DIGITS was applied to create the model for images classification.

In the DIGITS, before the network training hyparameters are able to be set by the user. Values of hypeparemeters, which were set during AlexNet and GoogLeNet training are presented in Table 1. For both networks almost all hypeparameters values are identical except the number of epochs. The number of epochs has significant influence for the training time and also for the training accuracy and training loss. The other hypeparametes, which are available to be set in the DIGITS are: base learning rate, batch size, snapshot interval, validation interval, random seed, batch accumulation.

Using the dataset described in the previous paragraph, the training of the network to create the model for gesture classification was done in the DIGITS, the results are presented on Fig. 3 and Fig. 4 and they depict relation between number of epochs versus accuracy, training loss and validation loss. Two different models were trained, one based on the AlexNet and the second one on the GoogLeNet network. The model based on the AlexNet architecture had maximum accuracy 98.87% at the epoch 24, and the second model, based on the GoogLeNet architecture had maximum accuracy 96.25% at the epoch 99.

The training was done on the workstation equipped with GPGPU card TESLA K20, which has 2496 computing cores. The workstation which works under Ubuntu 16.04 operating system. Multitude of cores results, that matrix operations (which are crucial during DNN training) were executed simultaneously and it accelerates computations significantly. To train the model with AlexNet network 2 minutes and 19 seconds were needed and 7.63 GB of the disk size space, in case of the GoogLeNet network 6 minutes and 9 seconds were needed and 3.88 GB of the disk size space. Both models were tested by loading to the DIGITS tool images from the hard disk and from the internet. The images used in tests were not applied previously for training neither for validation. Tests proved reliability of the models as they classified images correctly with likelihood equal even 99%.

Table 1. Hyperparameters values.

Hyperparameter	Values for AlexNet	Values for GooLeNet
Base learning rate	0.01	0.01
Batch size	10	10
Training epochs	35	100
Snapshot interval	1	1
Validation	1	1

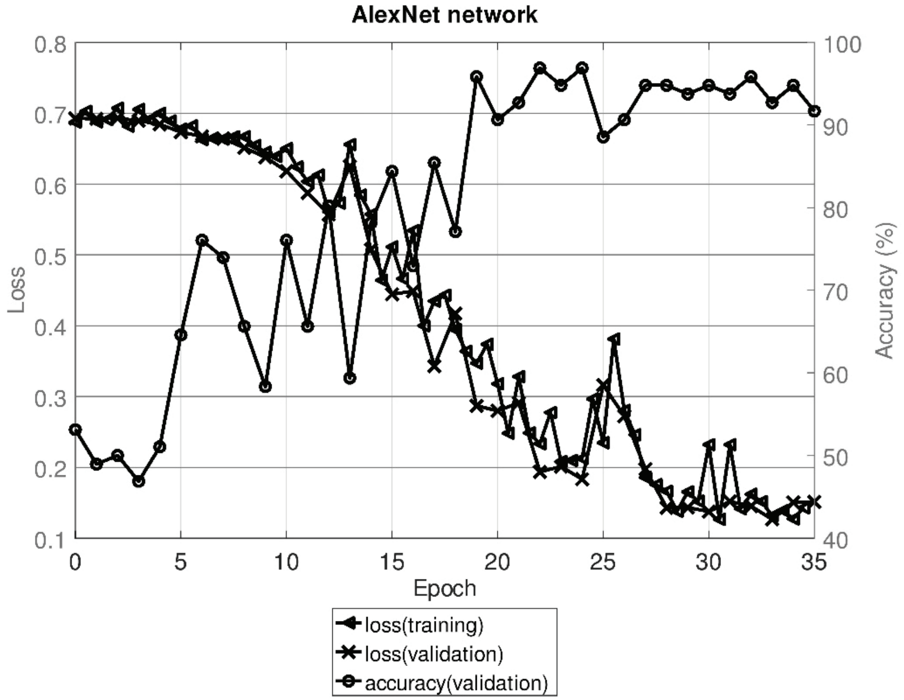

Fig. 3. Relations between number of epochs versus accuracy, training and validation losses, results obtained during the training of the AlexNet network

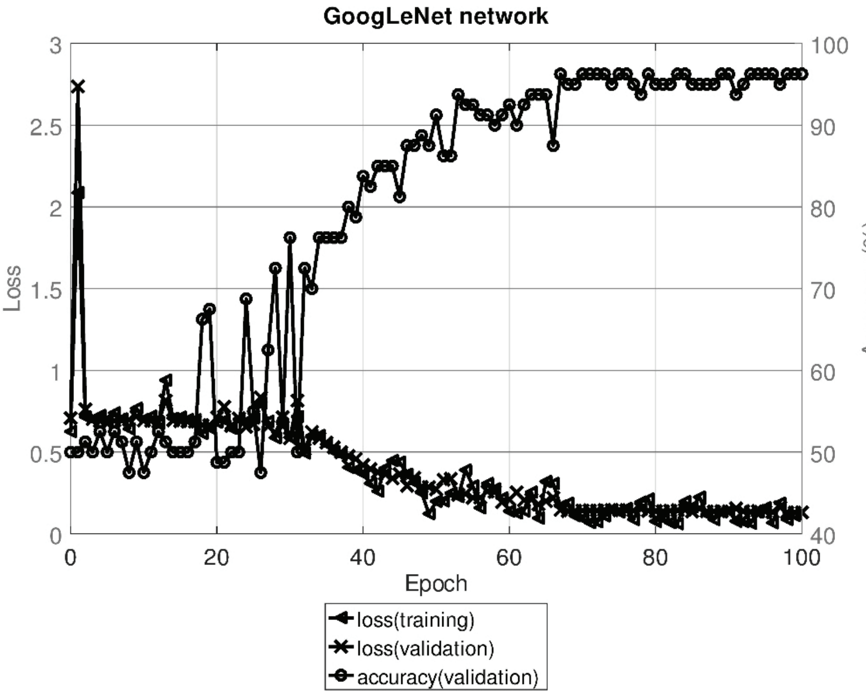

Fig. 4. Relations between number of epochs versus accuracy, training and validation losses, results obtained during the training of the GoogLeNet network

Both models were then deployed to the embedded GPU system called the development board Jetson TX1. Both models were empirically tested on the Jetson TX1, tests answered questions, if and how models recognize gestures of real time videos from the Jetson TX1 camera. The tests were executed with the use of the runtime library jetson-inference [10], which enables to run any code with the AI model compatible with the Caffe framework. Both models recognized two gestures correctly, however the model based on the AlexNet architecture gave the human user bigger gesture flexibility, the user could rotate hand with bigger number of angles than in case of the GoogLeNet. Due to that the model, based on the AlexNet architecture was chosen for further works.

4. FAN CONTROLLED BY GESTURE

Based on the model, described in the previous chapter the fan control system was created. The critical element of the system is the software, written in C and CUDA C programming languages. The software was embedded and executed on the Jetson TX1 development board. The software is based on runtime library jetson-inference and straightforward library to interface with GPIO (general-purpose input/output) pins [11]. The novelty of the paper is modification of the already existing code, which adjusts the software for requirements of the system with the fan.

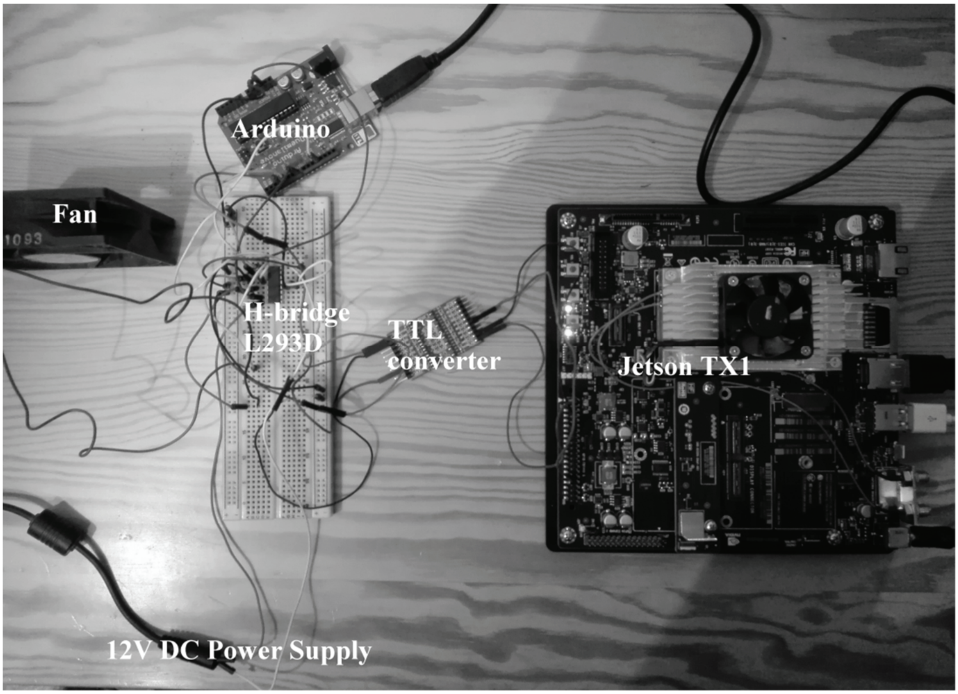

Fig. 5. Picture of the proposed prototype with the fan controlled visually by gestures

The dataflow, and data processing in the software is following: the camera of the Jetson TX1 board acquires, in real-time video images, then the image is classified, with the use of the model, created on the workstation, if the classifier will recognize "stone" gesture, then the software will set on the GPIO pin the logical value LOW, if the classifier will recognize "paper" gesture, then the software will set on the same GPIO pin the logical value HIGH.

Except the Jetson TX1 board, the system consists also TTL 8 Channel Logic Level Converter 3.3 V<->5 V, the Arduino Duemilanove board, based on ATmega328 microcontroller, L293D H-bridge motor driver integrated circuit, regulated DC power supply, 12 V fan, all elements of the system are presented on Fig. 5. The TTL converter is necessary as the GPIO interface applies 3.3 V digital logic and the rest of the system 5 V. The function of the Arduino board is to acquire signals of the GPIO Jetson TX1 interface and based on them to control the H-bridge L293D.

The control routine of L293D with the use of the Arduino board is following: if on the input digital pin number 3 of the Arduino board appears state HIGH then digital pins 4 and 8 are set respectively to the HIGH and LOW state and the analog

PWM (pulse width modulation) pin 10 is set to HIGH and duty cycle to 250, which is maximum value and results the maximum speed of the fan, in case on the input digital pin number 3 of the Arduino board appears state LOW then digital pins 4 and 8 are set like previously to the HIGH and LOW state and the analog PWM pin 10 is set to HIGH and duty cycle to 0, which is minimum value and results to stop the fan.

The integrated circuit L293D is applied to control inductive loads, including DC motors, the range of supply voltage is from 4.5 V up to 36 V. The fan of the created system is 12V DC motor. The outputs 1Y and 2Y of the L293D are connected to the fan directly, the L293D inputs 1,2EN, 1 A, 2 A are connected respectively to the pin 10, 4 and 8 of the Arduino board. The speed of the fan is controlled by PWM signal from pin 10 of the Arduino board, which enables or disables work of the L293D through the switching pin 1,2EN.

5. FAN CONTROLLED BY HANDWRITTEN NUMBERS

Beyond the system controlled by gestures, the second version of the system was created. This system enables to control the fan with the use of paper notes with handwritten digits. The design routine of the second system is fully analogous to the steps presented in chapter III and IV, due to that presented routines have universal character and enable to create a HMI system to control any HVAC with the use of any images, signs or shapes.

To control the fan four digits "0","1","2","3" were chosen, one of them, the digit “3” is presented on Fig. 6. The DNN model was trained based on the LaNet network [12], which was designed in the late eighties of the previous century. For the system, modifications which include four digits were made in the code and the classifier recognizes four cases instead of two. Also on the GPIO interface of the Jetson TX1 two pins were applied to encode four logical states. The encoding was following: if the classifier will recognize "0" then two GPIO pins will be set respectively to LOW and LOW logical states, if the classifier will recognize "1" then two GPIO pins will be set respectively to LOW and HIGH logical states, if the classifier will recognize "2" then two GPIO pins will be set respectively to HIGH and LOW logical states, , if the classifier will recognize "3" then two GPIO pins will be set respectively to HIGH and HIGH logical states. To receive encoded data from the GPIO interface also two digital pins were applied on the Arduino board. Through increase of the number of the fan controlling states, from two to four, the fan rotes with four speeds, maximum and minimum (which is zero) and two intermediate speeds. Similar to the system with gestures the system with handwritten digits works in real-time and is robust, reliable system.

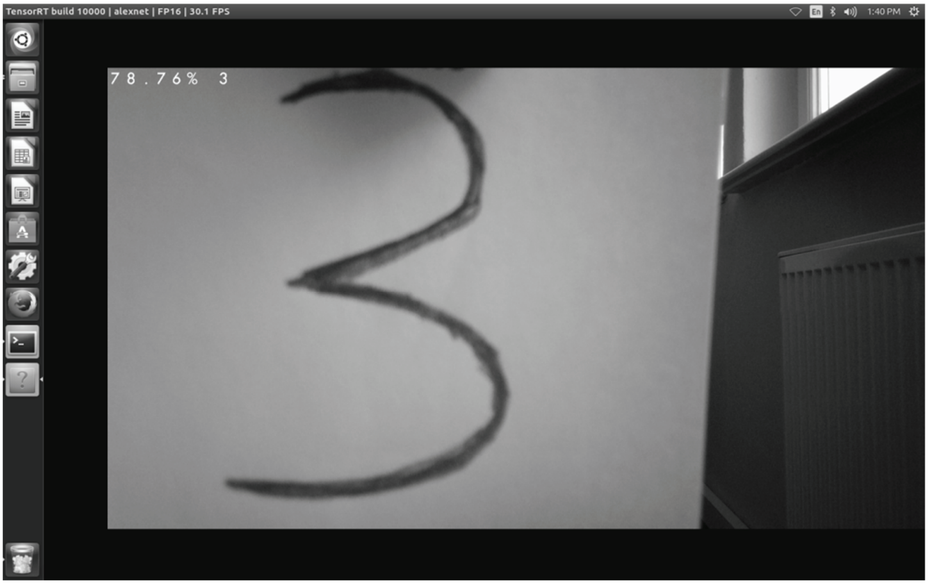

Fig. 6. Screenshot of the launched program to recognize paper notes with handwritten digits with the use of the Jetson TX1 camera

6. CONCLUSIONS

In the paper the vision based system to control the HVAC device with the use of gestures and handwritten digits was presented. The system relies on the embedded GPU supercomputer board Jetson TX1, which contains video camera and the board enables images processing and classification with the use of AI models, trained in CNN and DNN networks. To create AI models the Tesla K20, the DIGITS software and the LeNet, AlexNet, GoogLeNet networks were applied. Designed systems were tested and results validated, that the prototype systems are robust and enable unproblematic human machine interaction. The gesture based prototype can be potentially applied in autonomous "cars of the future", where passengers use gestures to control HVAC systems.

REFERENCES

[1] McCulloch W.S., Pitts W., A Logical Calculus of Ideas Immanent in Nervous Activity, Bulletin of Mathematical Biophysics, 5(4), pp. 115–133, 1943. https://doi.org/10.1007/BF02478259

[2] He K., Zhang X., Ren S., Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, In: IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1026–1034. https://doi.org/10.1109/ICCV.2015.123

[3] Arathi P. N., Arthika S., Ponmithra S., Srinivasan K., Rukkumani V., Gesture based home automation system, In: International Conference on Nextgen Electronic Technologies: Silicon to Software (ICNETS2), Chennai, India, 2017, pp. 198–201. https://doi.org/10.1109/ICNETS2.2017.8067929

[4] Ohn-Bar E., Trivedi M. M., Hand Gesture Recognition in Real Time for Automotive Interfaces: A Multimodal Vision-Based Approach and Evaluations, IEEE Tran. on Intelligent Transportation Systems, 15(6), pp. 2368–2377, 2014. https://doi.org/10.1109/TITS.2014.2337331

[5] Memo A., Minto L., Zanuttigh P., Exploiting Silhouette Descriptors and Synthetic Data for Hand Gesture Recognition, In: STAG: Smart Tools and Apps in computer Graphics, Verona, Italy, 2015, pp. 15–23. https://doi.org/10.2312/stag.20151288

[6] Memo A., Zanuttigh P., Head-mounted gesture controlled interface for human-computer interaction, Multimedia Tools & Applications, 77(1), pp. 27–53, 2018. https://doi.org/10.1007/s11042-016-4223-3

[7] Krizhevsky A, Sutskever I., Hinton G., ImageNet Classification with Deep Convolutional Neural Networks, In: 26th Annual Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, United States, 2012, pp. 1106–1114.

[8] Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A., Going deeper with convolutions, In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, United States, 2015, pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594

[9] NVIDIA, The NVIDIA Deep Learning GPU Training System, [online] Available at: https://developer.nvidia.com/digits [Accessed: 22 April 2019].

[10] NVIDIA, Guide to deploying deep-learning inference networks and deep vision primitives with TensorRT and NVIDIA Jetson, [online] Available at: https://github.com/dusty-nv/jetson-inference [Accessed: 22 April 2019].

[11] NVIDIA, A straightforward library to interface with the Jetson TX1 GPIO pins, [online] Available at: https://github.com/jetsonhacks/jetsonTX1GPIO [Accessed: 22 April 2019].

[12] LeCun Y., Bottou L., Bengio Y., Haffner P., Gradient-based learning applied to document recognition, Proceedings of the IEEE, 86(11), pp. 2278–2324, 1998. https://doi.org/10.1109/5.726791

(Received: 31.01.2020, revised: 09.03.2020)