USAGE OF BIO-INSPIRED NEURAL NETWORKS FOR THE DEAF SIGN SPEECH RECOGNITION[1]

Grif Mikhail Gennadievich
Doctor of Technical Sciences, Full Professor
Novosibirsk State Technical University, Novosibirsk, Russia
Kugaevskikh Alexander Vladimirovich
Candidate of Technical Sciences, Associate professor
Novosibirsk State University, Novosibirsk, Russia

Abstract. *The article offers an original approach to the recognition of sign speech of the deaf based on the use of a bio-inspired neural network. The main idea is to distinguish movement in sign language, which is typical for dynamic manual gestures and movements-epenthesis. The description of the motion detection model and its architecture is given. Motion detection is defined by the spatio-temporal scheme of the neural network organization. The first two layers select the edges. For this purpose, a neural network was developed based on the use of a Gabor filter and a hyperbolic tangent. The movement is directly isolated on the third layer of the neural network (MT-neuron). The MT-neuron allows you to distinguish both linear motion and rotation. The results of an experimental study of the motion detection model are presented, which confirm the expected hypotheses and its effectiveness for solving the problem of gesture speech recognition.*

Keywords: *sign language, gesture recognition, bio-inspired neural network, neuron movement.*

Introduction

There are about 150 national sign languages of the deaf in the world. The task of recognizing sign speech is relevant due to the lack of translators and their high cost. In practice, neural network approaches are used, but they also have their limitations [1]. Accurate recognition of sign speech is complicated by such phenomena of gesture display as dynamism, overlap, etc.The solution to this problem can be the use of motion recognition methods. For traditional neural networks, including those trained using deep learning, if the network is not trained for a certain motion vector, it will not detect it. In computer vision, the problem of

[1] The reported study was funded by RFBR and DST according to the research project No. 19-57-45006.

motion analysis is most often solved by applying the optical flow equation. When training neural networks to detect motion, we can also talk about using the optical flow equation, or rather, the basic mechanism for determining the direction of change in the brightness of pixels. Neural networks are also widely used in this task. Most often, reccurent networks, such as GRU [2, 3], LSTM [4, 5], are used to motiod detection. There is also a solution to the problem using convolutional neural networks [6-8], most often this is the ResNet model underlying FlowNet [9] and FlowNet 2.0 [10]. Several attempts have been made to simulate the motion detection neuron (MT-neuron) from the visual cortex. The most common model is the Heeger model [11, 12], which is analyzed in detail in [13]. In this paper, an alternative approach based on biological similarity is proposed.

Edge detection

In the visual cortex, motion analysis begins in the primary visual cortex. In the proposed model, the motion detection is carried out on the third layer of the neural network. The first two layers select the edges. For this purpose, a neural network was developed [14], based on the use of a Gabor filter and a hyperbolic tangent. The images coming to the input are represented by the L* component of the CIE space L*a*b*. On the first layer, lines of a certain orientation are detected. The second layer is responsible for selecting combinations of lines, including corners. Each layer contains 3 types of neurons that differ in the configuration of receptive fields. At the same time, the connections between the layers are organized in a special way. Each neuron of the second layer (U_{C2}) is connected only to two neurons of the first layer (U_{S1}). Thus, the neurons of the second layer allow you to detect lines and angles (in the case of the Gabor filter) and quadrilaterals (in the case of a hyperbolic tangent).

To lines detection, neurons are used, the receptive field of the size of 7*7 pixels of which is set by the Gabor filter.

$$G_{1,2} = exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda}+\varphi\right), \quad (1)$$

where $\frac{\sigma}{\lambda} \approx 0{,}56$, φ – phase offset (0 – dark line on a light background, $-\pi$ – light line on a dark background), γ – ellipticity (=0.1), $\lambda = 3$, $\theta \in [0{,}170]$.

Each neuron is sensitive to a line of a certain orientation with a deviation step of 10 degrees.

The neurons of the second layer are a weighted sum of the signals of the first layer with a sigmoidal activation function.

Motion detection

Motion detection sets the spatio-temporal organization of the motion detection neural network. Movement, in this case, is the sequential activation of several edge selection neurons

located in the same direction in a certain neighborhood over time, i.e. with a change of frame. Thus, the MT neuron can give the direction of movement α and its speed v. The MT neuron, like the previous neurons, is created for each type. The connections of the MT neuron with the U_{C2} neurons of the corresponding type determine its receptive field. To determine linear motion, the receptive field of the MT neuron ($U_{MT}^{\{l\}}$) includes a sequence of U_{C2} neurons in the α direction. To determine the rotation, the receptive field of the corresponding MT neuron ($U_{MT}^{\{r\}}$) is accompanied by connections with neurons located in the same center of the receptive field, but having different orientations θ. The rotation detection neuron is created twice for different directions of rotation.

$$U_{MT}^{\{l\}}(x,y,p,v,\alpha) = \sum_{x,y,t} U_{C2}(x,y,\theta,p) * w_{xy}(t) \quad (2)$$

$$U_{MT}^{\{r\}}(x,y,p,v,\alpha) = \sum_{\theta,t} U_{C2}(x_0,y_0,\theta,p) * w_{\theta}(t) \quad (3)$$

The weights of MT-neurons are set using the product of Gaussian and Mexican Hat wavelet, the first is responsible for the spatial characteristic, the second sets the attenuation coefficient of the link weight over time.

$$w_{xy}(t) = exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right) * exp\left(1-\frac{t^2}{2\sigma^2}\right) exp\left(-\frac{t^2}{2\sigma^2}\right) \quad (4)$$

$$w_{\theta}(t) = exp\left(-\frac{\theta^2}{2\sigma^2}\right) * exp\left(1-\frac{t^2}{2\sigma^2}\right) exp\left(-\frac{t^2}{2\sigma^2}\right) \quad (5)$$

A uniform filling of such a neuron, i.e. a stationary dark area in the entire size of the receptive field, will not give the required activation, fig.1. In this case, the attenuation coefficient obeys a certain law of change t: at the beginning of the movement in the receptive field $t = [0,1,2]$, when the neuron $U_{C2}(x_2,y_2,\alpha,p)$ is activated, the vector t will have values $t = [-1,0,1]$. At the end of the receptive field, the vector t will have values $t = [-2,1,0]$. In this regard, the value of the attenuation coefficient will change.

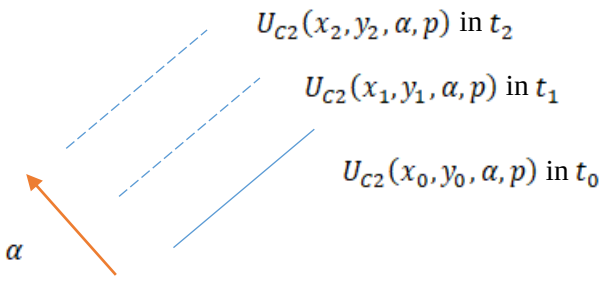

Fig. 1. Scheme of the MT neuron work

The Mexican hat wavelet is used to reduce activation when filling the receptive field of the MT neuron with a textural periodic image.

Experiments

For an experimental test, we will run the movements of different angles to check the activation of U_{C2} neurons, in the direction of 45 degrees. Fig. 2-4 show the frame-by-frame activation of MT neurons. Ideally, there should be a thin line, but due to the low resolution, there is a false activation within 10-20 degrees ($\sigma = 0.5$).

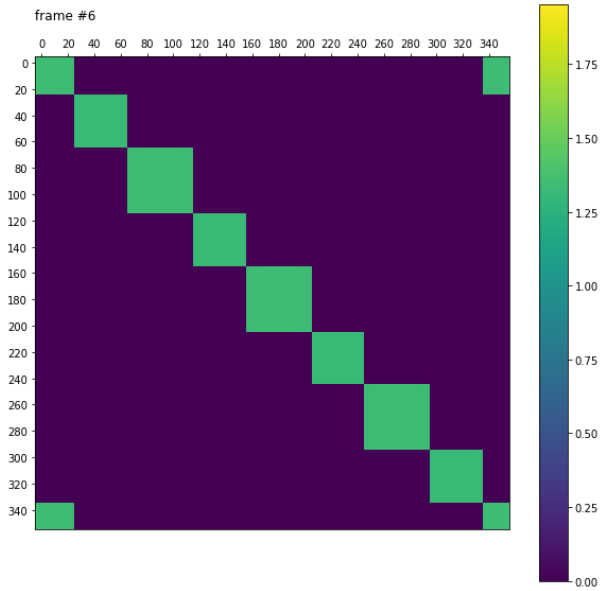

Fig. 2. First frame (*start motion, vertically - the angle of movement, horizontally-the response of MT-neurons*)

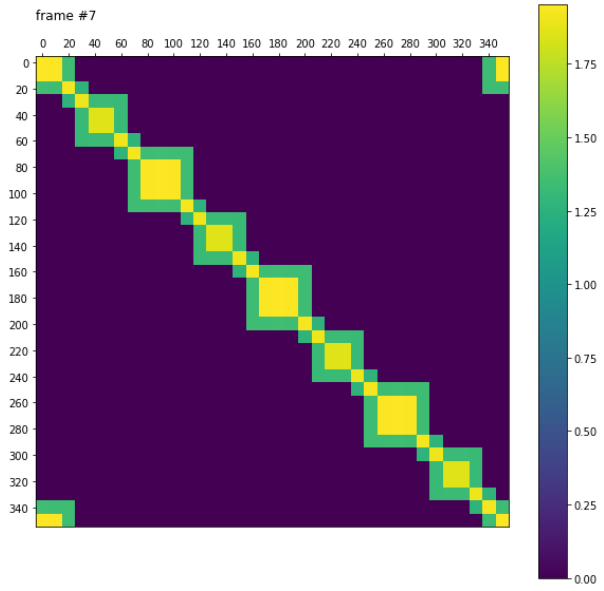

Fig. 3. Second frame (*vertically - the angle of movement, horizontally-the response of MT-neurons*)

Fig. 4 shows that the maximum activation is achieved on the second frame, then there is a fade, which confirms our assumption. Attenuation is necessary so that the MT neuron does not fire on stationary objects.

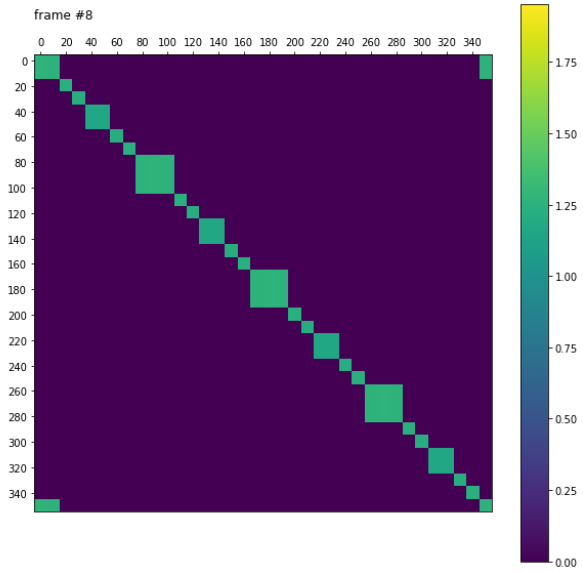

Fig. 4. Third frame (*end motion, vertically - the angle of movement, horizontally-the response of MT-neurons*)

In general, the longer the movement lasts, the more accurately its direction is determined.

If we increase the size of the receptive field of the space-time vector from 3 to 7, the accuracy of determining the direction of movement increases, fig.5-6.

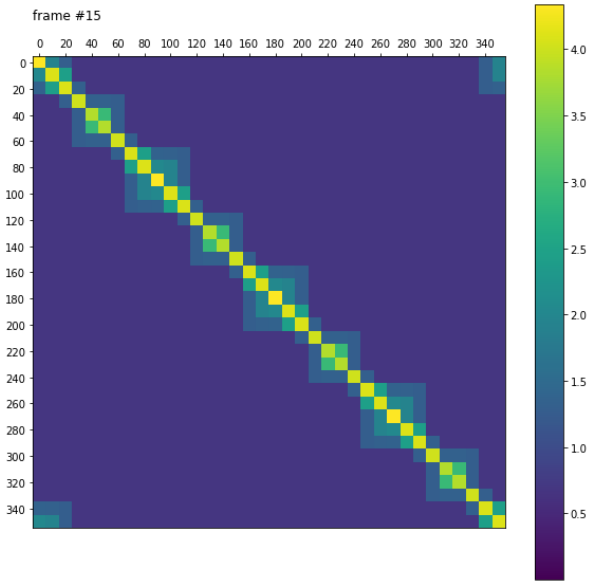

Fig. 5. Fourth frame (vertically - the angle of movement, horizontally-the response of MT-neurons)

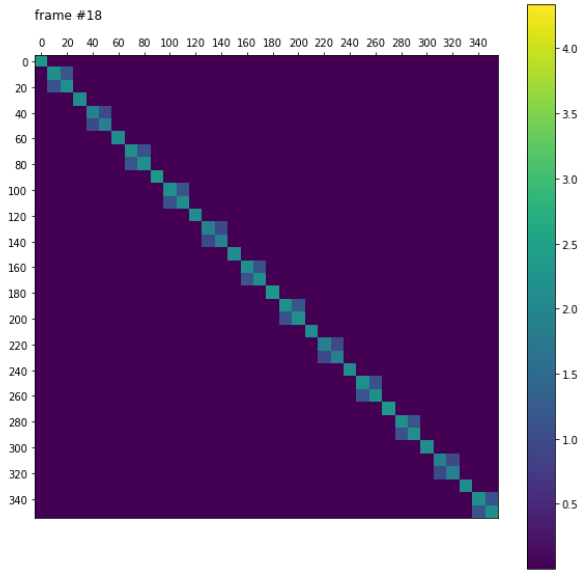

Fig 6. Seventh frame (*end motion, vertically - the angle of movement, horizontally-the response of MT-neurons*)

Conclusions

To recognize the sign language of the deaf, it is proposed to use a biologically similar neural network. The basis for its application is the presence of dynamic manual gestures and movements in sign language-epenthesis. Motion detection is defined by the spatio-temporal scheme of the neural network organization. The first two layers select the edges based on the use of the Gabor filter and the hyperbolic tangent. The movement is directly isolated on the third layer of the neural network (MT-neuron). The MT-neuron allows you to distinguish both linear motion and rotation. The experimental study of the motion detection model confirmed the expected hypotheses and its effectiveness for solving the problem of gesture speech recognition.

References

1. *Prikhodko A.L., Grif M.G., Bakaev M.A.* Sign language recognition based on notations and neural networks // *Communications in Computer and Information Science*. - 2020. - Vol. 1242: DTGS 2020: Digital Transformation and Global Society. - pp. 463-478. - DOI: 10.1007/978-3-030-65218-0_34.

2. Cai Y., Liu J., Guo Y., Hu S. and Lang S. Video anomaly detection with multi-scale feature and temporal information fusion //*Neurocomputing*, Vol. 423, pp. 264–273, Jan. 2021, DOI: 10.1016/j.neucom.2020.10.044.

3. *Tokmakov P., Schmid C. and Alahari K.* Learning to Segment Moving Objects // *International Journal of Computer Vision*, Vol. 127, No. 3, pp. 282–301, Mar. 2019, DOI: 10.1007/s11263-018-1122-2.

4. *Szeto R., Sun X., Lu K. and Corso J.J.* A Temporally-Aware Interpolation Network for Video Frame Inpainting //IEEE Trans. Pattern Anal. Mach. Intell., Vol. 42, No. 5, pp. 1053–1068, May 2020, DOI: 10.1109/TPAMI.2019.2951667.

5. *Pavllo D., Feichtenhofer C., Auli M. and Grangier D.* Modeling Human Motion with Quaternion-Based Neural Networks // *Int J Comput Vis*, Vol. 128, No. 4, pp. 855–872, Apr. 2020, DOI: 10.1007/s11263-019-01245-6.

6. *Kim D., Woo S., Lee J.Y. and Kweon I.S.* Recurrent Temporal Aggregation Framework for Deep Video Inpainting // *IEEE Trans. Pattern Anal*. Mach. Intell., Vol. 42, No. 5, pp. 1038–1052, May 2020, DOI: 10.1109/TPAMI.2019.2958083.

7. *Wang H., Yu B., Xia K., Li J. and Zuo X*. Skeleton edge motion networks for human action recognition //*Neurocomputing*, Vol. 423, pp. 1–12, Jan. 2021, DOI: 10.1016/j.neucom.2020.10.037.

8. *Maczyta L., Bouthemy P. and Le Meur O.* CNN-based temporal detection of motion saliency in videos // *Pattern Recognition Letters*, Vol. 128, pp. 298–305, Dec. 2019, DOI: 10.1016/j.patrec.2019.09.016.

9. *Dosovitskiy A. et al.* FlowNet: Learning Optical Flow with Convolutional Networks // IEEE International Conference on Computer Vision (ICCV), Santiago, Dec. 2015, pp. 2758–2766. DOI: 10.1109/ICCV.2015.316.

10. *Ilg E., Mayer N., Saikia T., Keuper M., Dosovitskiy A. and Brox T.* FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, Jul. 2017, pp. 1647–1655. DOI: 10.1109/CVPR.2017.179.

11. *Heeger D.J.* Model for the extraction of image flow // *J Opt Soc Am A*, Vol. 4, No. 8, pp. 1455–1471, Aug. 1987, DOI: 10.1364/josaa.4.001455.

12. *Simoncelli E.P. and Heeger D.J.* A model of neuronal responses in visual area MT // *Vision Research*, Vol. 38, No. 5, pp. 743–761, Mar. 1998, DOI: 10.1016/S0042-6989(97)00183-1.

13. *Chessa M., Sabatini S.P. and Solari F.* A systematic analysis of a V1–MT neural model for motion estimation // *Neurocomputing*, Vol. 173, pp. 1811–1823, Jan. 2016, DOI: 10.1016/j.neucom.2015.08.091.

14. *Kugaevskikh A.V. and Sogreshilin A.A.* Analyzing the Efficiency of Segment Boundary Detection Using Neural Networks // *Optoelectron.Instrument.Proc.*, Vol. 55, No. 4, Jul. 2019, DOI: 10.3103/S8756699019040137.