A VISUAL COMMUNICATION MODEL BY NEURAL NETWORKS

Assoc. prof. Todor Kostadinov, PhD*, Zhanina Dubarova-Kostadinova, PhD**
**Burgas Free University **National academy of arts*

Abstract: *Visual communication, rooted in the fundamental human activity of drawing, offers universal accessibility across linguistic and cultural barriers. A neural network-based model to interpret sequential drawings as a languages, enabling communication between speakers of different languages, aiding interaction of autistic children is proposed. The article integrates visual language concepts with artificial intelligence, describing a neural network design. Applicable in multilingual communication and assistive technologies are explored.*

Key words: *Visual communication, neural networks, multi-language communication, linguistic model.*

МОДЕЛ НА ВИЗУАЛНО ОБЩУВАНЕ С НЕВРОННИ МРЕЖИ

доц. д-р Тодор Костадинов*, д-р Жанина Дубарова-Костадинова**
**Бургаски Свободен Университет **Национална художествена академия*

Резюме: *Визуалната комуникация, основана на фундаменталната човешка дейност на рисуването, предлага универсална достъпност, преодолявайки езиковите и културни бариери. Предложен е модел, базиран на невронна мрежа, който интерпретира последователности от рисунки в писмен език, позволявайки комуникация между говорещи различни езици, деца с аутизъм в изразяването на техните мисли и емоции. Статията интегрира концепции от визуалния език и изкуствения интелект, описвайки за целта модел на невронна мрежа. Приложим в многоезична комуникация и подпомагащи технологии.*

Ключови думи: Визуална комуникация, невронни мрежи, многоезично общуване, *лингвистичен модел.*

1. INTRODUCTION

Even before the children learn to speak, they are learning about visual objects and shapes. Then the knowledge of shapes and images is bounded to sounds and later to letters and symbols. Before the appearance of literacy, the main communication was in verbal form and if the language becomes a barrier then no communication can be made. The purpose of literacy is mainly to preserve the knowledge. The main disadvantage is that it is not universal for the mankind. However, the gestural language can be taken as a middle part between literacy and images. It means also that the process of communication initially uses images. Drawing, one of the earliest forms of human expression, has served as a universal communication medium throughout history. Its ability to transcend linguistic and cultural barriers is unparalleled, offering a direct and intuitive way to convey ideas and emotions.

Drawing can be interpreted as a visual language with its own set of symbols and structures, enabling individuals to express complex concepts without relying on spoken or

written words [1]. The evolution of visual communication highlights its effectiveness in transmitting complex ideas rapidly. From early cave paintings to modern visual design, the principles underlying this form of interaction remain consistent: simplicity, universality, and immediacy It has been shown in [2] how deep learning architectures can mimic human visual perception, bridging the gap between natural and artificial intelligence. The potential of convolutional neural networks CNNs in recognizing and interpreting visual patterns is demonstrated in [3]. In recent decades, advancements in artificial intelligence AI have provided tools to bridge traditional visual communication with modern computational methods. Long Short-Term Memory LSTM networks, for instance, excel in understanding temporal sequences, making them suitable for tasks requiring contextual interpretation [4]. These developments present a unique opportunity to reinterpret drawing as a structured and translatable form of communication, suitable for diverse applications ranging from multilingual interaction to assistive technologies for children with autism. This study seeks to align the ancient practice of drawing with contemporary AI techniques. This work introduces a novel method for intelligent visual communication, by employing neural networks to interpret drawings as structured linguistic outputs. This method could facilitate multilingual interaction, where drawings are interpreted and translated into sentences in specified languages, and assistive communication for individuals who face challenges with verbal expression [5]. The increasing demand for inclusive and versatile communication technologies underscores the significance of this approach. In contexts such as education, therapy, and creative industries, the ability to convert visual inputs into semantic outputs could transform how people interact with technology and each other. Faster R-CNN, which enables real-time object detection and classification, enhancing the model's ability to handle complex visual data in real-time applications [8]. Similarly, in natural language processing (NLP) [9] underlines how visual data and language can be intertwined through machine learning, creating possibilities for deeper communication systems. Further, a contribution to the understanding of large-scale visual recognition challenges with ImageNet, providing a robust framework for training deep learning models to recognize complex visual inputs has been made in [10]. [11] outlined the theoretical foundations for understanding and training neural networks, providing the basis for integrating visual data with textual interpretation. Additionally, insights into how CNNs process visual data hierarchically, which is essential for understanding the role of image processing in the proposed model is made in [12]. This study builds upon these foundational works in visual cognition [1], neural network design [3], and human-computer interaction [8], proposing a comprehensive framework for integrating drawing.

2. VISUAL COMMUNICATION AND ARTIFICIAL INTELLIGENCE

The intersection of visual communication and AI is an exciting and rapidly evolving area of research. For centuries, visual communication, particularly through drawing and other forms of imagery, has been essential in human interaction. Unlike verbal language, which is confined by grammatical and syntactical rules specific to each language, visual communication transcends linguistic barriers. This makes it a powerful tool in conveying complex ideas, emotions, and concepts in a universally comprehensible format. A drawing operates as a universal language, using symbols and images to communicate messages that are understood by individuals from diverse cultural backgrounds. Historically, visual communication has played an important role in conveying meaning across various societies. Early forms of visual language, such as cave paintings and hieroglyphs, paved the way for the symbolic representation of thoughts and actions. In modern times, visual media,

from digital art to graphic design, continues to bridge the gap between complex concepts and human understanding.

However, as technology has advanced, the ability to translate these visual messages into a format that can be processed by machines has become increasingly critical. This is where artificial intelligence comes in, particularly through its application in image recognition and natural language processing - NLP. AI technologies, especially deep learning algorithms, have demonstrated remarkable success in mimicking human visual perception and pattern recognition. Convolutional Neural Networks - CNNs are at the heart of this progress. CNNs are designed to replicate the way the human brain processes visual information. They use layers of convolutional filters to extract spatial hierarchies from images, allowing machines to identify patterns, objects, and features within visual data. These networks have been trained on vast datasets, such as ImageNet, which contains millions of labeled images, enabling them to recognize everything from simple shapes to complex objects in images [9], [10]. This ability is particularly relevant to visual communication, where the goal is to interpret visual inputs (e.g., drawings or images) and map them to meaningful outputs (e.g., sentences or semantic descriptions).Furthermore, AI's role in image recognition extends beyond just identifying objects; it also involves understanding the context and temporal relationships between visual elements. This becomes especially important in interpreting sequential images, such as a series of drawings. Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), have proven effective in capturing these temporal dependencies. By analyzing the relationships between drawings in a sequence, LSTM networks can understand the progression of a visual narrative, similar to how humans interpret a story told through images or gestures. This ability is fundamental in the proposed neural network model for interpreting sequences of drawings as structured linguistic outputs. The integration of CNNs for feature extraction and LSTMs for sequence interpretation creates a hybrid model capable of transforming visual input into coherent text output. Artificial intelligence, in particular deep learning and computer vision techniques, thus offers a bridge between the visual and linguistic worlds. By processing visual data, AI can help us interpret not just what is visible, but also what it means, providing a deeper understanding of the images that convey information. This approach is pivotal for creating systems that can automatically translate visual symbols into language, breaking down barriers between speakers of different languages and enhancing communication across diverse contexts. As neural networks evolve, their potential applications extend far beyond traditional image recognition tasks. The integration of visual communication with AI also opens doors to innovative solutions in fields such as assistive technologies, education, and interactive platforms for artistic expression. For instance, AI-driven tools can aid children with autism by interpreting their drawings and turning them into speech or text, providing an outlet for expression that they might otherwise struggle to access [8]. Similarly, AI can help break down the communication barriers in multilingual settings by converting drawings into text that can then be translated into multiple languages.

In drawing as a Language, images can convey complex ideas to people from diverse cultures quickly and effectively Fig. 1.

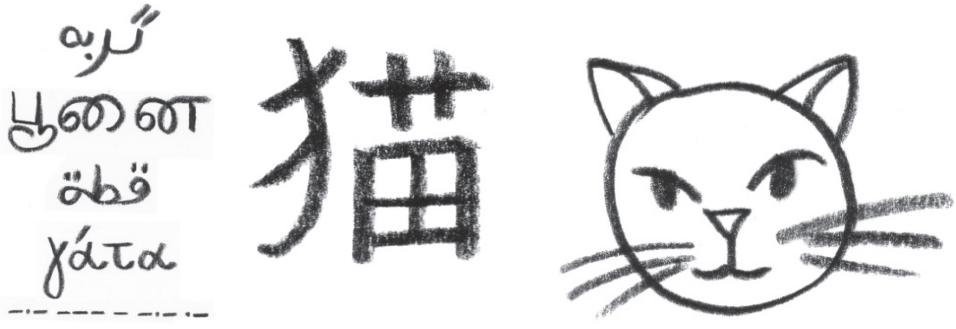

Figure 1. Relation between text and drawings

The combination of visual communication and artificial intelligence represents a promising area for advancing both human interaction and machine learning. By leveraging the power of AI, particularly CNNs and LSTMs, to decode and interpret drawings, we can create intelligent systems that not only understand visual input but also generate meaningful linguistic output. This has the potential to revolutionize how humans communicate across languages and abilities, offering a new, inclusive approach to interaction that transcends the limitations of traditional verbal and written language.

3. MODEL OF THE NEURAL NETWORK

As a graphically related task it is necessary to choose the appropriate AI tools to deal with images. For that purpose, convolutional neural networks – CNN are mainly used for image recognition and classification. Other features of the CNN is the extraction of objects, shapes, edges, and textures. These features are progressively refined and transformed into high-level representations in the deeper layers. If interpretation of sequential data such as multiple images, the LSTM layers capture the temporal relationships between the images in the sequence. The LSTM units retain information about previous drawings, ensuring that the model understands how individual drawings relate to each other in context. Finally, the fully connected and output layers can generate text from the learned representations. This could be a single word, a phrase, or a more complex sentence, depending on the desired output.

The proposed model employs a CNN architecture. Input layers process image sequences representing drawings, which are transformed into feature maps using convolutional layers. Subsequent recurrent layers, such as Long Short-Term Memory (LSTM) units, interpret temporal dependencies between drawings, akin to understanding word sequences in a sentence. The network's output layer generates text in the specified language.

The architecture of the neural network presented in Fig. 2 is comprised of five layers:

- Input Layer – Accepting grayscale images of drawings,
- Convolutional Layers – Extracting spatial features,
- LSTM Layers – Capturing sequential dependencies,
- Output Layer – Providing translated textual output.

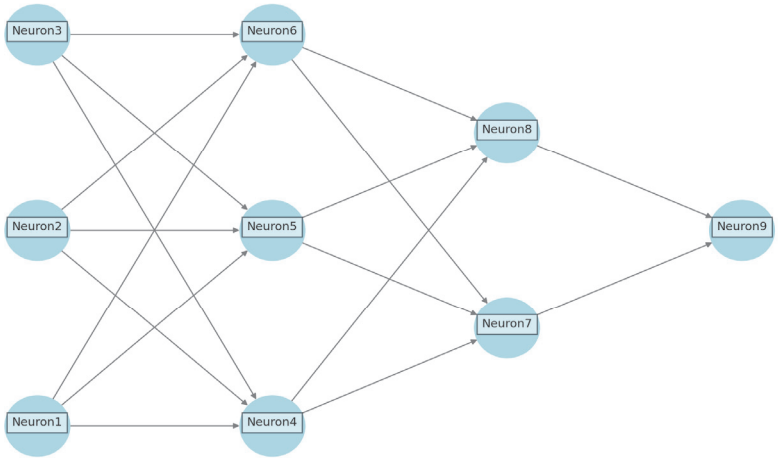

Figure 2. Neural Network model for Visual Communication

According to Fig. 1, the neurons interconnection helps to process visual input data through various layers of feature extraction to generate meaningful output. In that sense Neuron1, Neuron2, Neuron3 form the Input Layer. It receives grayscale images, and each neuron processes one pixel of the drawing. These neurons represent the input, which corresponds to the pixels of the drawing. The convolutional layer 1 is comprised of Neuron4, Neuron5, Neuron6. Where spatial features from the input images, such as edges, shapes, and textures, crucial for recognizing elements in the drawing are extracted. The second convolutional layer - Neuron7, Neuron8 (Convolutional Layer 2). In this layer, more complex features are extracted, such as interactions between objects or finer details that help in better understanding the drawing. Neuron9 is in the Output Layer. It generates the final textual output based on the processed drawing, reflecting the generated text that describes or translates the drawing into the specified language.

The proposed neural network model is designed for visual communication by processing sequential drawings into textual language outputs. Below is a description of the layers and their connections. In the context of training a neural network for interpreting sequential images such as drawings and provide words or semantic text output, the dataset should consist of images, their associated labels - words, and the derived text output. The input images are the primary data fed into the neural network for processing. These images represent drawings or sketches that encode visual information. For this type of system, the images should be of appropriate size and resolution. Each image should be resized to a consistent size, typically a square 128×128 for best performance. Also the training vectors with grayscale images are smaller, that the RGB images. To represent drawings or sketches the images must be line drawings, sketches, or image illustrations as training Vectors, the neural network needs both the input image and a corresponding label that provides the correct output. Each image is associated with a vector or label that contains the desired text or semantic meaning. The simplest form of labeling involves associating each drawing with a word or a phrase. In the training phase, the neural network needs to be provided with pairs or sequences of images and the corresponding output text. If the network is trained on sequential data, the corresponding text should match the sequence, ensuring that the network learns the correlation between image elements and linguistic components.

During training, the neural network must be provided with image-label pairs. The network learns to map visual features from the images to linguistic representations (words, phrases, sentences). The network's loss function measures how close its generated text is to the target text, and the model adjusts its weights using backpropagation to minimize this loss.

CONCLUSION

This study presents a novel approach to visual communication using neural networks, making the relation of drawings and linguistic output. By applying convolutional layers for spatial feature extraction and LSTM layers for sequential dependencies, the proposed model interprets visual data and translates it into meaningful text. This integration of drawing, an ancient and universal form of communication, with artificial intelligence technologies has the potential to revolutionize how people interact across linguistic, cultural, and cognitive boundaries. The applications of this model vary from multilingual communication, assistive technologies for individuals with disabilities, to interactive educational tools. The ability to generate semantic textual interpretations from sequential drawings opens new opportunities for inclusive communication, particularly for children with autism or individuals facing challenges with verbal expression and disabled. Future research will focus on improving the ability of the neural network adaptability to distinguish drawing styles and enhancing its robustness to noisy inputs. Additionally, exploring larger, more complex datasets can be used for model validation. Expanding the system to handle interactive feedback and context-aware drawing sequences could make it even more effective in educational and therapeutic contexts.

REFERENCES

[1] Жанина Дубарова, „РИСУВАНЕТО КАТО ЕЗИК”, Unpublished manuscript, 2023.
[2] Y. LeCun, Y. Bengio, and G. Hinton, „Deep learning”, Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[3] A. Krizhevsky, I. Sutskever, and G. Hinton, „ImageNet classification with deep convolutional neural networks,“ Advances in Neural Information Processing Systems, vol. 25, pp. 1097-1105, 2012.
[4] S. Hochreiter and J. Schmidhuber, „Long short-term memory,“ Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[5] K. He, X. Zhang, S. Ren, and J. Sun, „Deep residual learning for image recognition“, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, 2016.
[6] R. Girshick, „Fast R-CNN“, Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448, 2015.
[7] M. Abadi et al., „TensorFlow: A system for large-scale machine learning“, Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, pp. 265–283, 2016.
[8] S. Ren, K. He, R. Girshick, and J. Sun, „Faster R-CNN: Towards real-time object detection with region proposal networks“, Advances in Neural Information Processing Systems, vol. 28, pp. 91–99, 2015.
[9] R. Collobert et al., „Natural language processing (almost) from scratch“, Journal of Machine Learning Research, vol. 12, pp. 2493–2537, 2011.
[10] O. Russakovsky et al., „ImageNet large scale visual recognition challenge“, International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015.
[11] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.
[12] M. Zeiler and R. Fergus, „Visualizing and understanding convolutional networks“, Proceedings of the European Conference on Computer Vision, pp. 818–833, 2014.