Ткач Микола Олександрович

Київський національний університет технологій та дизайну

(м. Київ)

Науковий керівник – Ренська І. І.

THE CAPABILITY OF MICROSOFT AI TO CREATE OBJECTS BASED ON DETAILED TEXT DESCRIPTIONS

If your task is to draw a picture of a bird with a yellow body, black wings and a short beak, then most likely you will start with the coarse contour of the birds, as this is the basis, then look at the note, seeing the yellow part and take the yellow handle, To fill the body, read the note again and take a black pen to draw the wings, and in the end, cut the beak and determine its reflective shine. Then you can draw a branch of a tree on which the bird sits.

Now there is a bot that can do it.

The new artificial intelligence technology developed by Microsoft research laboratories is programmed to focus on specific words when creating images from text descriptions of signatures.

The bottle closes the research circle around the intersection of computer vision and the processing of natural language, which he and his colleagues investigated during the last half-decade. They started with a technology that automatically captures CaptionBot captions and then moves to technology that answers questions about people asking for an image, such as a location or attributes of objects that can be especially useful for blind people and for people with disabilities, as well as for children. [4]

To do this, you need to teach modelling machines for identifying objects, interpreting actions and exchanging natural language.

Image production is a more complicated task than image captions, an associate researcher in a team; because the process requires that the bot image presents parts that are not contained in the title. "This means that you need your

machine learning algorithms, starting your artificial intelligence to present some inaccessible parts of the images." [2]

At the heart of technology, the bot for Microsoft is a technology called Generative Adversarial Network or GAN. The network consists of two models of machine learning, one of which generates images from text descriptions, and another, known as a discriminator, which uses text descriptions to determine the authenticity of generated images. The generator receives false pictures relating to the discriminator; the discriminator does so not to be deceived. Working together, the discriminator clicks the generator on the perfection and accuracy of the transmission of the generated image by text description.

Microsoft's Microsoft bot studied on data sets that contain paired images and signatures that allow models to learn to pick words up to the visual representation of these words. For example, GAN learns to generate bird image when the title is called a bird, and also find out what bird picture should look like. "This is a fundamental reason why we believe that a car can be trained." [3]

The GAN works well when creating images from simple text descriptions such as blue bird or evergreen trees, but the quality of stagnation consists of more complex detailed text descriptions, such as a bird with a green crown, yellow wings, and a red abdomen. This is because the whole sentence serves as the only input of the generator. Detailed description information is lost. As a result, the generated image is a blurred greenish-yellowish-red bird instead of a close, sharp coincidence with the description.

The technology of creating text-to-image technology can find practical applications that work as an original assistant to artists and interior designers, or as a tool for activating voice processing of photos. With more computational power, he imagines that technology can generate by extracting certain hand-made animated films based on scripts, reinforcing the work performed by animated filmmakers.

Today's technology is imperfect. Approximate examination of images almost always detects defects, for example, birds with blue bees, and not black

and fruit plants with mutant bananas. These disadvantages are an obvious indication that the computer, and not the person, created the image. However, the quality of the AttnGAN image is almost three times the improvement compared to the previous GAN's best in class and, as it thinks, serves as a milestone in the overall, human analysis that raises human capabilities. [1]

Thanks to such technologies it is possible to make life easier. The computer will perform amazing transformation for you in a graphic image from a text description. For the greatest accuracy of such a procedure, improvement is needed in order to avoid problems with the mistaken perception of some information.

REFERENCES

1. John Roach "Microsoft researchers build a bot that draws what you tell it to". [Electronic resource]: [Web site]. – Electronic data. – Access mode: https://blogs.microsoft.com/ai/drawing-ai/ – Title of the screen

2. Rob LeFebvre "Microsoft AI can draw objects based on detailed text descriptions". [Electronic resource]: [Web site]. – Electronic data. – Access mode: https://www.engadget.com/2018/01/18/Microsoft-ai-draws-images-from-detailed-text/ – Title of the screen

3. Sarah Perez "Microsoft's new drawing bot is an AI artist". [Electronic resource]: [Web site]. – Electronic data. – Access mode: https://techcrunch.com/2018/01/18/microsofts-new-drawing-bot-is-an-a-i-artist/ – Title of the screen

4. Pedro Hernandez "Microsoft's AI-Powered Drawing Bot Turns Text into Images". [Electronic resource]: [Web site]. – Electronic data. – Access mode: http://www.eweek.com/enterprise-apps/microsoft-s-ai-powered-drawing-bot-turns-text-into-images – Title of the screen