V. Savinskyi, Student
V. Spivachuk, PhD in Engr, As. Prof., research advisor
Khmelnytsky National University

THE IMPACT OF AI ON THE WORLD OF INFORMATION PROCESSING

The aim of this study was to show resent studies in computer science field named artificial intelligence and show recent researches, that have been made in the past year. Considered methods can help in different areas — from the military industry to video production industry.

My talk in about AI — artificial intelligence. Today we live in a world where we are surrounded by a lot of computer per square meter. As for me, I see big power and big interest in this, because those computers don't just exists around us — they are getting information about us.

There small computers, that you all have in your pocket, is actually a listening device plus camera. This is impactful not only for us — humans with natural intelligence — but also for AI.

Today AI is relevant to any intellectual task. Modern artificial intelligence techniques are pervasive and are too numerous to talk about them. Frequently, when a technique reaches mainstream use, it is no longer considered artificial intelligence; this phenomenon is described as the AI effect.

Let me tell you about 3 recent researches that AI scientists have made in the resent past.

This paper [1] describes a new technique to visualize the inner workings of a generator neural network. This is a neural network that is is able to create images for us. The key idea here is dissecting this neural network, and looking for agreements between a set of neurons and concepts in the output image, such as trees, sky, clouds, and more. This means, analyzing that some neurons are responsible for buildings to appear in the image, and those will generate clouds.

Interestingly, such agreements can be found, which means way more than just creating a visualizations, that is shown in paper, because it enables us to edit images without any artistic skills.

The editing part works by forcefully activating and deactivating these units and correspond to adding or removing these objects from an image. This means that we can take an already existing image, and ask this technique to remove trees from it, or perhaps add more, the same with domes, doors, and more.

This is already awesome, but there is more. Note that so far, the amount of control we have over the image is quite limited. And fortunately, we can take this further, and select a region of the image where we wish to add something new. This is suddenly so much more granular and useful!

The algorithm seems to understand that trees need to be rooted somewhere and not just appear from thin air. Most of the time anyway. Interestingly, it also understands that bricks don’t really belong here, but if I add it to the side of the building, it continues it in a way that is consistent with its appearance. Most of the time anyway.

Most research works are but a step in a thousand-mile journey, and each of them tries to improve upon the previous paper. This means that a few more papers down the line, this will probably take place in HD, perhaps in real-time, and with much higher

quality. This work also builds on previous knowledge on "generative adversarial networks", and whatever the follow-up papers will contain, they will build on knowledge that was found in this work.

You are free to improve this, because the authors kindly made the source code available free for everyone, and not only that, but there is also a web app so you can also try it yourself! This is an excellent way of maximizing the impact of your research work. I have a feeling that lots of high-quality entertainment materials will surface very soon.

Next AI Senses Humans Through Walls.

Pose estimation is an interesting area of research where we typically have a few images or video footage of humans, and we try to automatically extract the pose a person was taking. In short, the input is one or more photo, and the output is typically a skeleton of the person. So what is this good for? A lot of things. For instance, we can use these skeletons to cheaply transfer the gestures of a human onto a virtual character, fall detection for the elderly, analyzing the motion of athletes, and many others.

This work [2] showcases a neural network that measures how the wifi radio signals bounce around in the room and reflect off of the human body, and from these murky waves, it estimates where we are. Not only that, but it also accurate enough to tell us our pose. As the wifi signal also traverses in the dark, this pose estimation works really well in poor lighting conditions. That is a remarkable feat. But, there is more. We know, that wifi signals go through walls. So perhaps, this means that we can see through walls? And this, what was done in this work. It tracks the pose of this human as he enters the room, and, as he disappears, the algorithm still knows where he is. This means that it can also detect our pose through walls! What kind of wizardry is that?

Now, note that this technique doesn't look at the video feed we are now looking at. It is there for us for visual reference. It is also quite remarkable that the signal being sent out is a thousand times weaker than an actual wifi signal, and it also can detect multiple humans. This is not much of a problem with color images, because we can clearly see everyone in an image, but the radio signals are more difficult to read when they reflect off of multiple bodies in the scene.

The whole technique work through using a «teacher-student network» structure. The teacher is a standard pose estimation neural network that looks at a color image and predicts the pose of the humans therein. So far, so good, nothing new here. However, there is a «student network»б that looks at the correct decisions of the teacher, but has the radio signal as an input instead. As a result, it will learn what the different radio signal distributions mean and how they relate to human positions and poses. As the name says, the teacher shows the student neural network the correct results, and the student learns how to produce them from radio signals instead of images.

If anyone said that, they were working on this problem ten years ago, they would have likely ended up in an asylum. Today, it's reality.

Next AI Learned To Isolate Speech Signals.

This is [3] a neural network-based technique that can perform audio-visual separation. Before we talk about what that is, I will tell you what it is not.

This new technique can clean up an audio signal by suppressing the noise in a busy bar, even if the source of the noise is not seen in the video. It can also enhance the voice of the speaker at the same time.

Alternatively, if we have a skype-meeting with someone in a lab or a busy office where multiple people are speaking nearby, we can also perform a similar speech separation, which would be a godsend for future meetings.

Moreover, I think if you are a parent, the utility this example needs no further explanation. I am not sure if I ever encountered the term "screaming children" in the abstract of an AI paper, so that one was also a first here.

This is a super difficult task, because the AI needs to understand what lip motions correspond to what kind of sounds, which is different for all kinds of languages, age groups, and head positions.

To this end, the authors put together a stupendously large dataset where he have almost 300.000 videos with clean speech signals. This dataset is then run through a multi-stream neural network that detects the number of human faces within the video, generates small thumbnails of them, and observes how they move over time.

It also analyzes the audio signal separately, and then fuses these elements together with a recurrent neural network to output the separated audio waveforms. A key advantage of this architecture and training method is that as opposed to many previous works, this is speaker-independent, therefore we do not need specific training data from the speaker we want to use this on. This is a huge leap in terms of usability.

The paper also contains an excellent demonstration of this concept by taking a piece of footage from Conan O'Brien's show where two comedians were booked for the same time slot and talk over each other.

The result is a performance where it is near impossible to understand what they are saying, but with this technique, we can hear both of them one by one, crystal clear.

REFERENCES

1. Inbar Mosseri Ariel Ephrat. “Looking to Listen at the Cocktail Party: A Speaker-Independent Audio-Visual Model for Speech Separation". In: (2019).

2. David Bau et al. “GAN Dissection: Visualizing and Understanding Generative Adversarial Networks". In: Proceedings of the International Conference on Learning Representations (ICLR). 2019.

3. Tianhong Li Mingmin Zhao. “Through-Wall Human Pose Estimation Using Radio Signals". In: (2019).