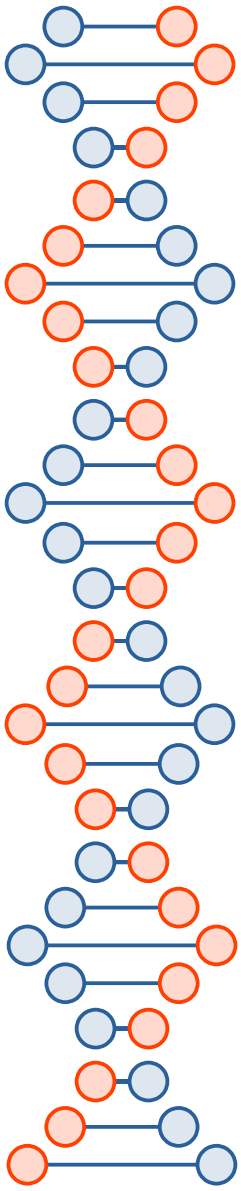
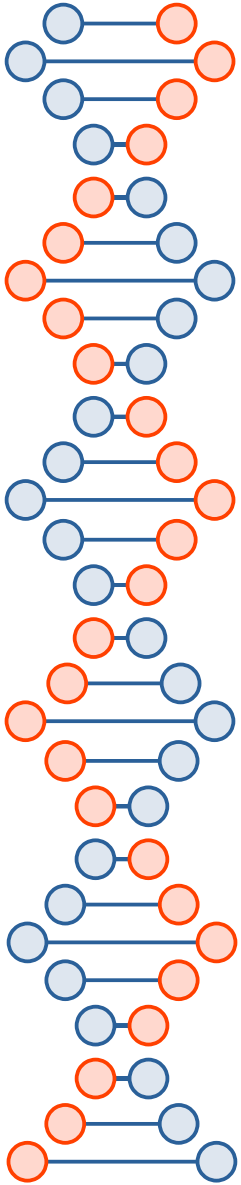


JETSON AI CERTIFICATIONS

Yet Another Virtual Room

Jetson AI Ambassador



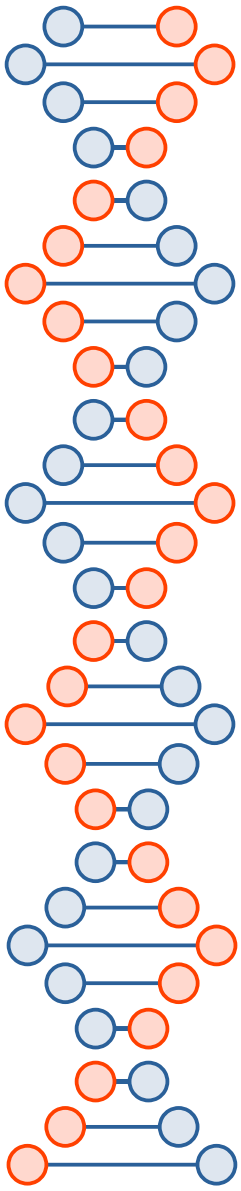
JETSON AI CERTIFICATIONS

About me.

I have been a senior lecturer at Southern Federal University (Russia, Rostov-na-Donu) for many years and have taught cources of many subjects including "Machine Learning Technology".

In addition, I teach the AI course at the Samsung IT Academy https://myitacademy.ru/ai/ai_lectures/

Currently I am writing a Ph.D. thesis on AI / ML

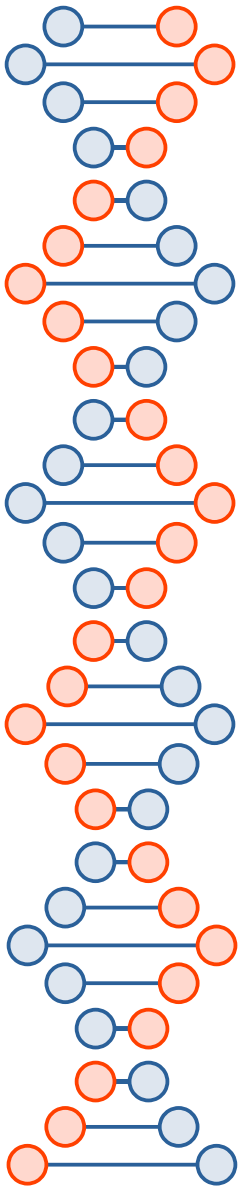


JETSON AI CERTIFICATIONS

Idea.

I wanted to develop a practically useful application that I myself and anyone else could use. In my teaching career, I need the ability to replace the background when communicating through remote communication programs. Unfortunately, not all programs can do this. For example MC Teams or Discord do not implement this functionality.

Therefore, I decided to get useful value from the purchase of Yatson nano and, as a project, write an application that does this.

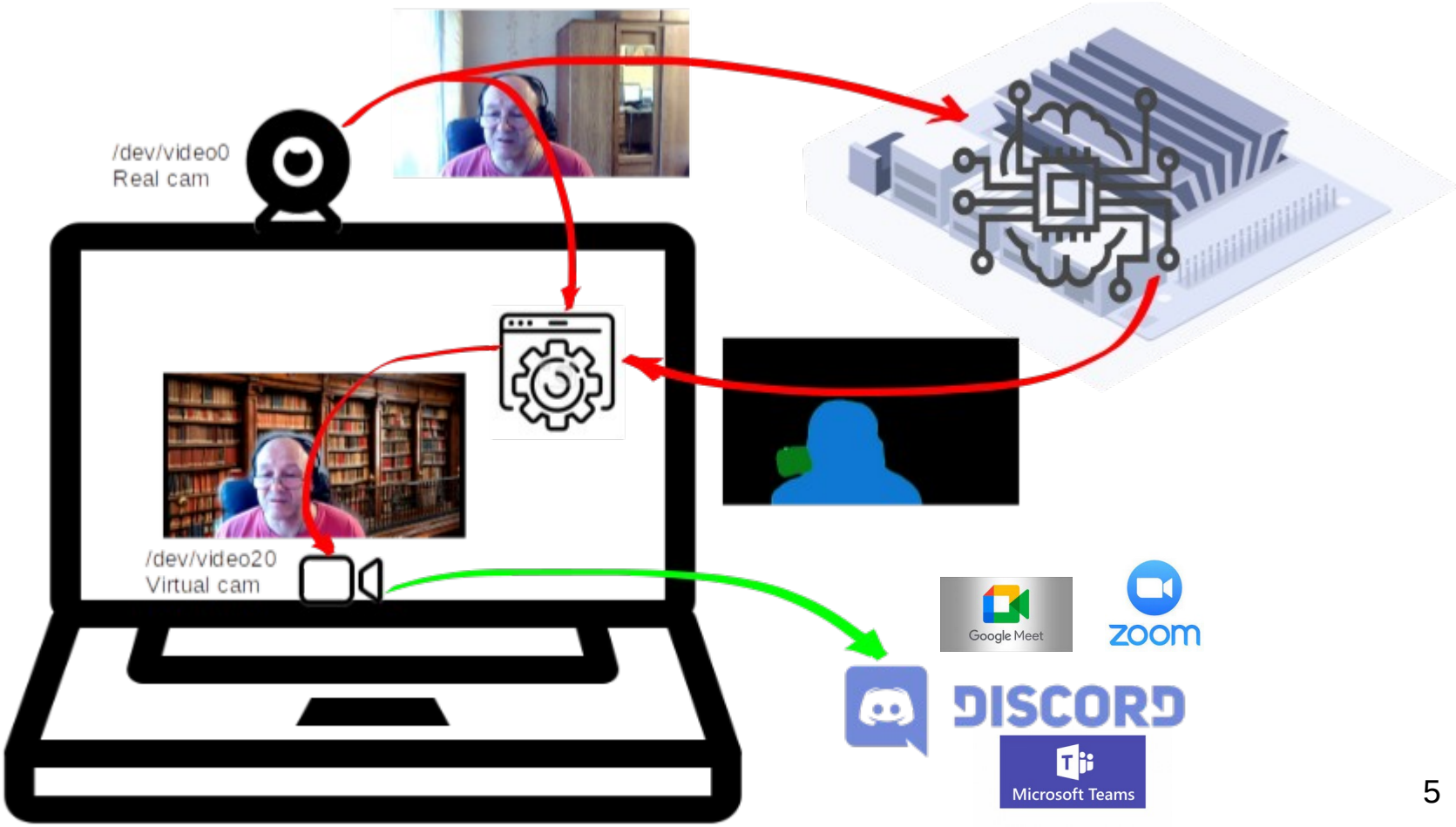


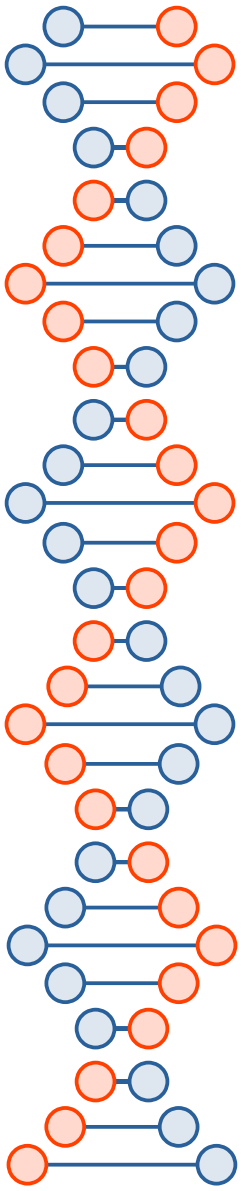
JETSON AI CERTIFICATIONS

Tasks.

1) Capture images from webcam to image series
2) Semantic segmentation object from image
3) Create mask for image by colors
4) Generate new image by masked images from (1)
5) Put image from (4) to virtual (aka fake,looback and so on) video device v4l2loopback
6) Configure remote communcation software to use loopback video device as my video camera

JETSON AI CERTIFICATIONS





JETSON AI CERTIFICATIONS

Challenges.

- My laptop, like most laptops, is too slow for streaming segmentation - you need to transfer processing to another device
- For the normal functioning of the program, the prediction speed must be less than 50 ms. It's too fast for Jatson - optimization needed
- Due to the distributed way of working, you need the largest possible asynchronous and fault-tolerant program code

JETSON AI CERTIFICATIONS

Hardware and software.

- Laptop (client)
  - Dell 5491 i5-8300H 2.3Gh/2*8G/SSD256
  - OpenSUSE Leap 15.3, 5.3.18-59.19-default
  - Python 3.6.13 on bare metal
- Jetson Nano (server)
  - Jetson B01 NANO 2G
  - Ubuntu 18.04.6 LTS, 4.9.201-tegra
  - Python 3.6.9 on bare metal

JETSON AI CERTIFICATIONS

Client implementation.

The client side app consists of two modules ***ui.py*** and ***main.py*** . Source code https://github.com/d-yacenko/YAVR/tree/main/client

ui.py is nothing interesting - it's just a graphical user interface written in tkinter, working with a config file and call with the selected parameters main.py

main.py performs more important tasks

- loads the v4l2loopback module (you need to configure sudoers for the current user)
- frame-by-frame capture from a web camera and transmission of frames via an http post request to a web server on Jetson for segmentation. Here you need to limit the request in time to achieve real-time work, and handle the possibility of the server not providing a result
- retrieves back a mask image from the server
- apply a mask to remove background, apply static or dynamic background, apply special effects

JETSON AI CERTIFICATIONS

Server implementation.

Source code - https://github.com/d-yacenko/YAVR/tree/main/server . The server side of the application consists of two modules **server.py** and **segmentation.py**

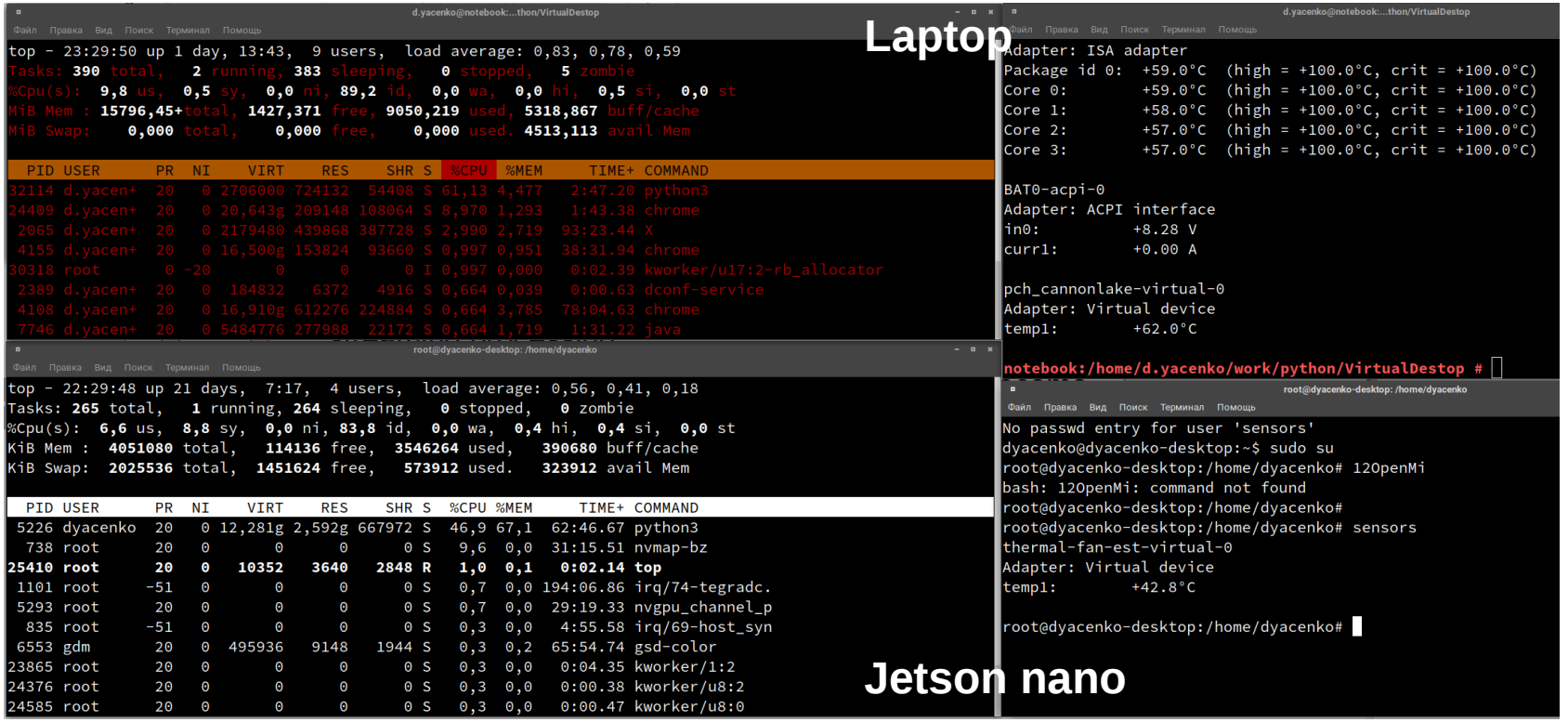
- **server.py** is a web server implementation where the input image is received in the post request and run segmentation.py module to process image it and get the mask. This is where threads are used for optimization.
- **segmentation.py** is the most interesting module in the whole project. The pretrained network deeplabv3_mobilenet_v3_large from the torchvision project was selected for segmentation. This network showed the highest segmentation speed rate. I decided not to train my own network, because I would hardly have gotten a higher speed, and it is speed that is critically important in the application. In addition, a number of other techniques have been undertaken to speed up the inference:
  - Using non-blocking pinned memory to speed up GPU → CPU data transfer
  - Reducing the bitness of data (half) before transferring to the CPU
  - Computing argmax also on GPU
  - Cleaning up memory on GPU after inference
  - Note. For a number of reasons, it was necessary to save the input image and the generated mask as a file. A ramdisk was used to speed up I/O and protecting storage from being burned out. To mount it, an entry was added to /etc/fstab: ***tmpfs /RAM tmpfs size = 10m 0 0***

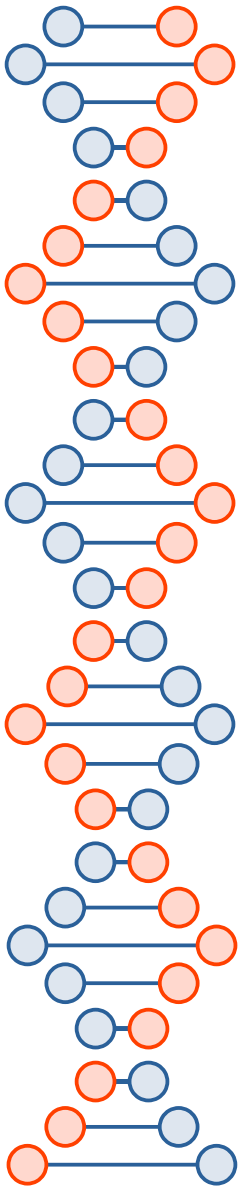
JETSON AI CERTIFICATIONS

Results.

- MVP application was created for low cost real-time streaming processing.
- Source code of application - https://github.com/d-yacenko/YAVR.git
- Video demo is here: https://drive.google.com/file/d/1VlaSLGa5P0YtwAdNDjHJI9W7D7Tcl3PS/view?usp=sharing
- The application is useful and has value and can become a startup
- The load on the equipment and the temperature of the components during the operation of the application are not high (see next slide)

JETSON AI CERTIFICATIONS

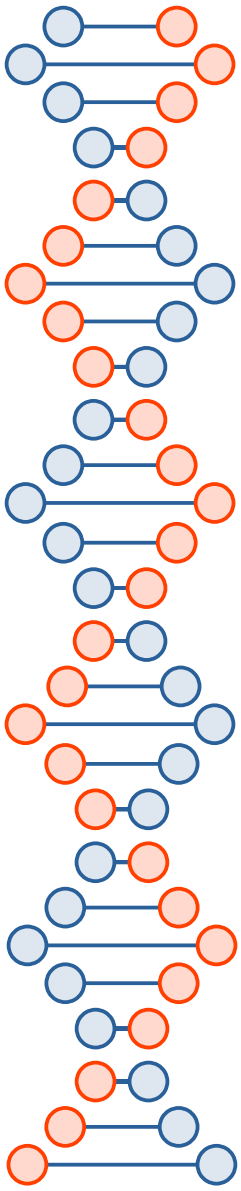




JETSON AI CERTIFICATIONS

Conclusions.

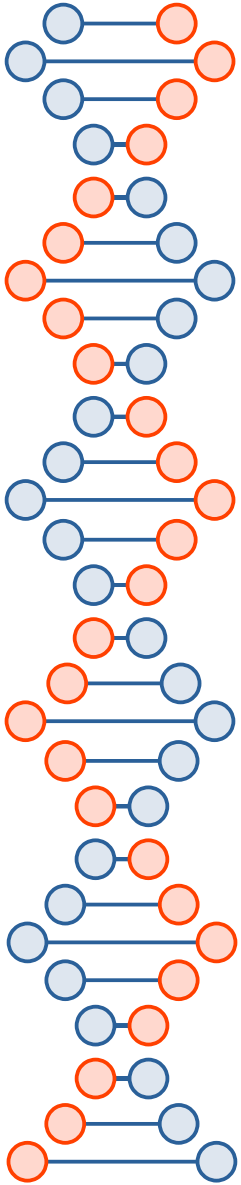
- Jetson Nano level devices can be used for not very large real AI/ML tasks
- In part, the Jetson Nano can work as an edge device, for example, at the junction of IoT and AI.



JETSON AI CERTIFICATIONS

Motivation.

- Why did I start doing this project? It became interesting for me to explore the capabilities of the Jetson nano and my AI/ML capabilities on specialized equipment.
- Why am I getting an Ambassador Certification? First of all, as a teacher, a certificate will be useful to me as a confirmation of status. Secondly, maybe the Ambassador program will give me the opportunity to get other Nvidia equipment for research - for example, the Xavier NX, or some other.



Thanks for attention!