

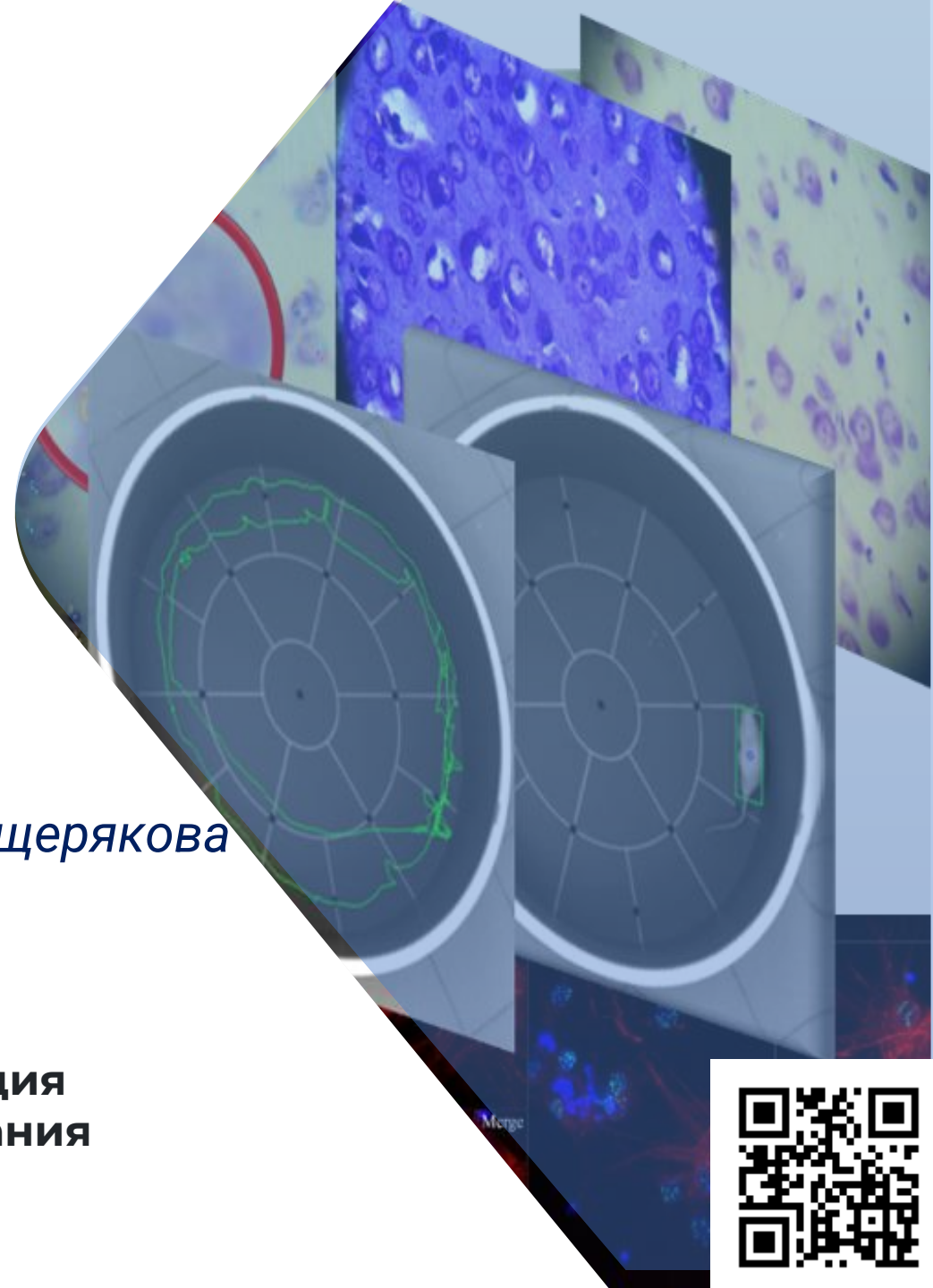
Сервисы информационной системы на основе методов машинного обучения и технологий анализа данных для совместного проекта BIONLIT (ЛИТ и ЛРБ ОИЯИ)

О.И. Стрельцова

Лаборатория информационных технологий им. М.Г. Мещерякова
Объединенный институт ядерных исследований
От имени группы **BIONLIT**

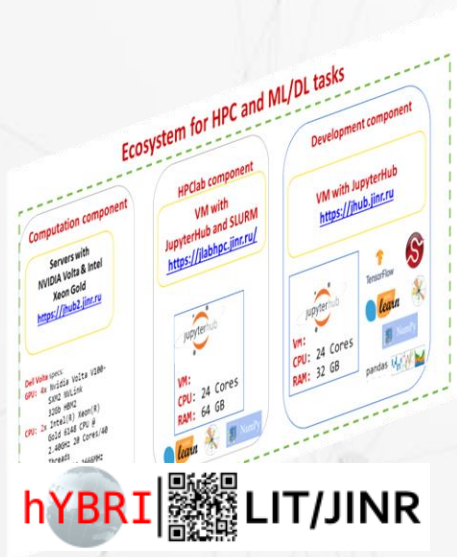
**V Всероссийская научно-практическая конференция
«Перспективы развития математического образования
в эпоху цифровой трансформации»**

г. Тверь, 28 – 30 марта 2024 года





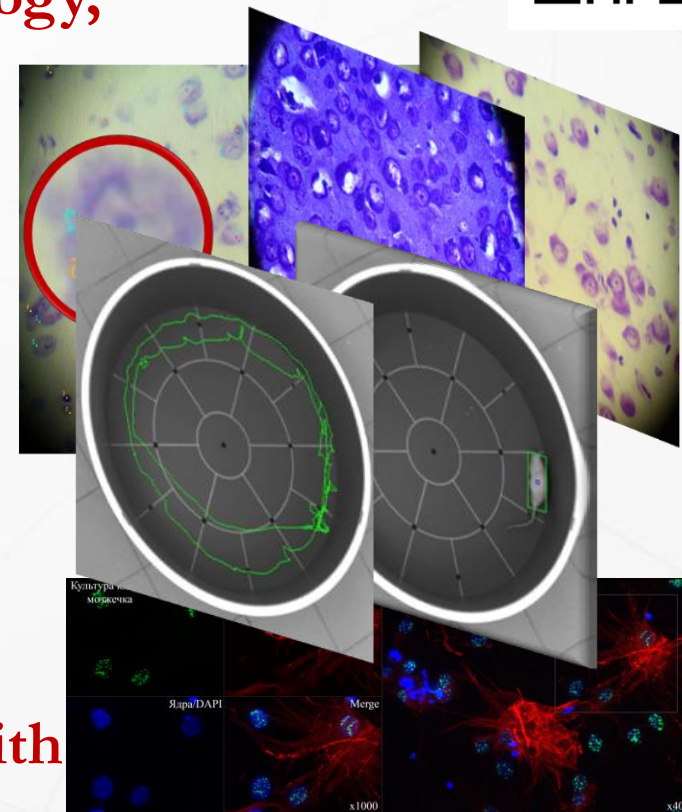
Meshcheryakov Laboratory of Information Technologies, JINR



Anikina A.I.
 Bezhanyan T.Zh.
 Nechaevskiy A.V.
 Podgainy D.V.
 Shadmehri S.
 Streltsova O.I.
 Zuev M.I.

Laboratory of Radiation Biology, JINR

Boreyko A.V.
 Kolesnikova I.A.
 Lyakhova K.N.
 Severiukhin Yu.S.
 Utina D.M.
 Chausov V.N.
 Chramko T.S.



With the participation of undergraduate and graduate students

Dubna University, TSU, RUDN University

In collaboration with Stadnik A.V.

In collaboration with In collaboration with



Coordinator: Dr Marko Ćosić

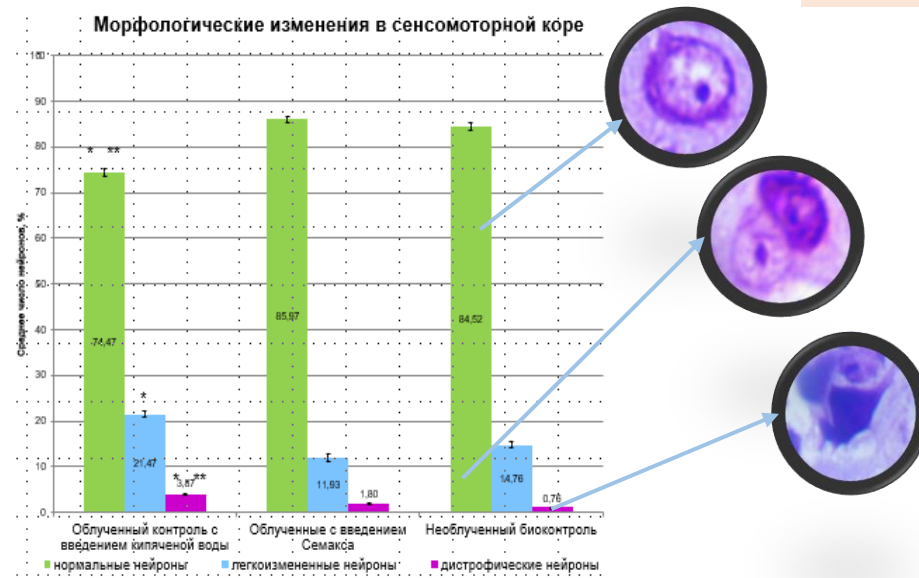
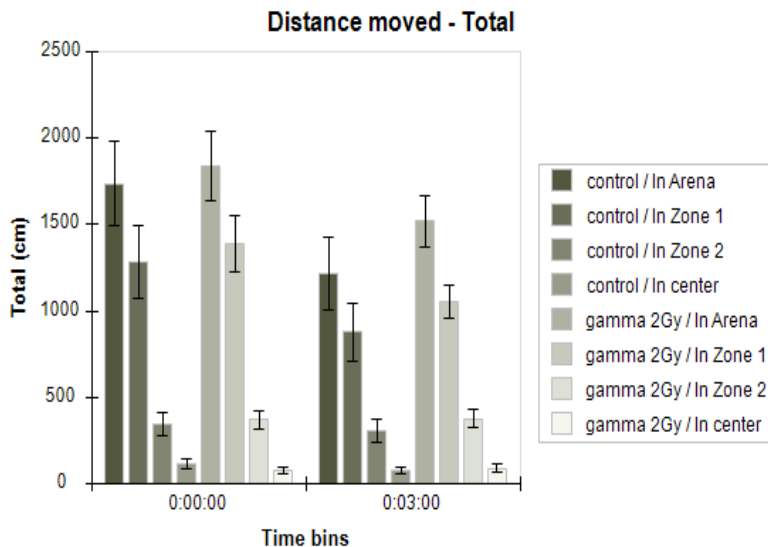
Germany:
 Streltsov Alexei I.
 Gromov E.



Stage #1

Комплексный анализ полученных результатов на всех этапах исследования

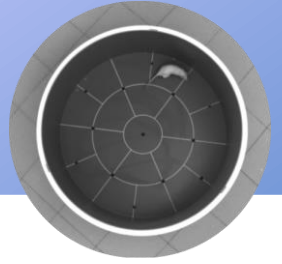
Stage #2



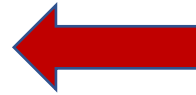
ИС необходима для:

- **Хранения** всего массива экспериментальных данных (фото- и видеоматериалы, файлы pdf, excel, doc с информацией об эксперименте) с целью минимизации риска потери информации.
- **Обеспечения** удобного доступа к данным для всех членов исследовательской группы и обеспечения возможности проводить как пошаговый, так и комплексный анализ данных в едином информационном пространстве.
- **Сокращение** времени, затрачиваемого на обработку данных, увеличение скорости получения качественных результатов и снижение субъективности подхода к обработке экспериментальных данных.

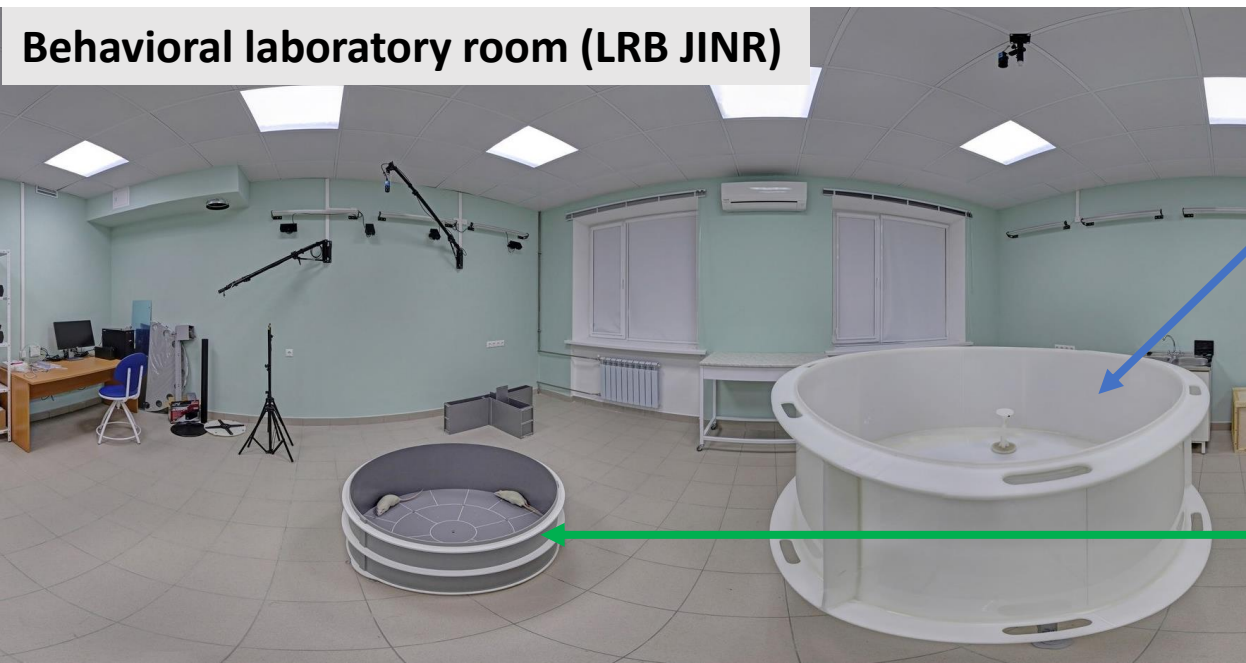
Stage #1: Анализ поведенческих реакций лабораторных животных



Первый этап связан с анализом поведенческих реакций лабораторных животных, подвергшихся воздействию ионизирующего излучения, а также возможного влияния на поведение фармакологических препаратов.



С точки зрения анализа данных этот этап связан со сбором и анализом **видеоданных**, отражающих поведенческие реакции лабораторных животных в ходе экспериментальных испытаний на специализированных стендах.



Behavioral laboratory room (LRB JINR)

Test system “**Water Maze**”

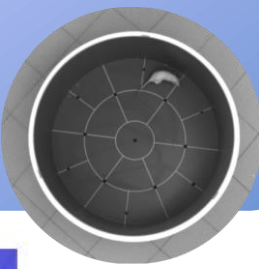
Test system “**T-maze**”



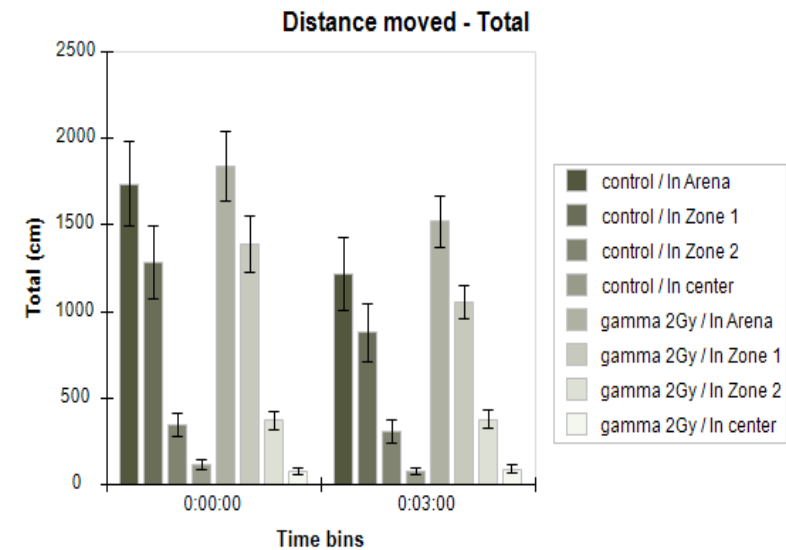
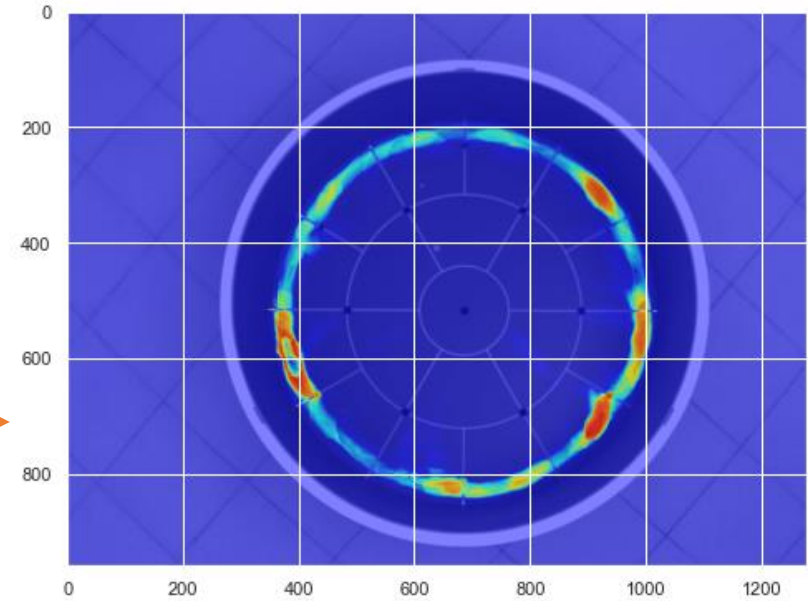
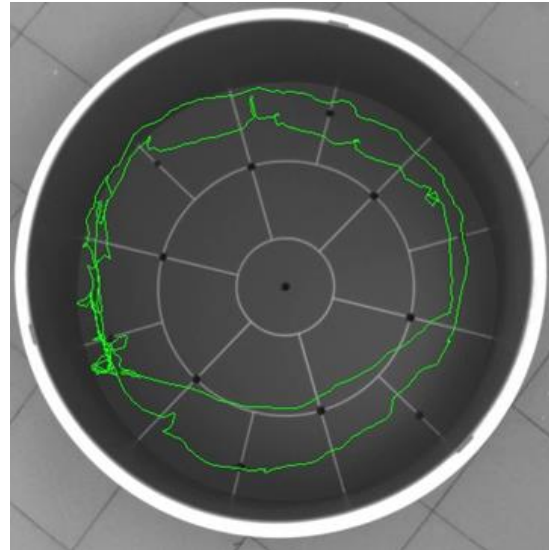
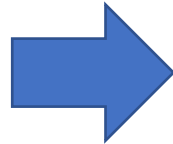
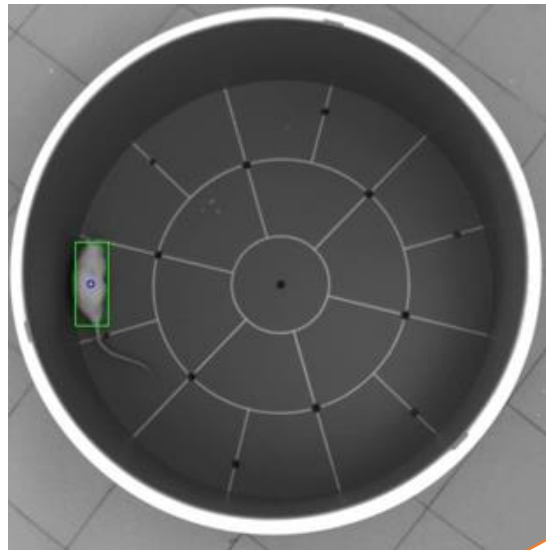
Test system “**Open Field**”

“**Beam-walking**” (SPC Open Science, Russia)

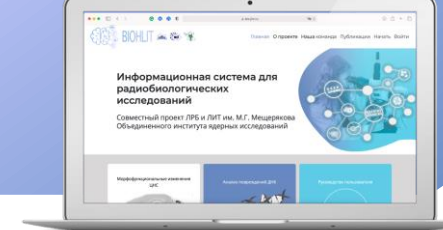
Stage #1: Анализ поведенческих реакций лабораторных животных



Test system "Open Field"



ML/DL/computer vision algorithms

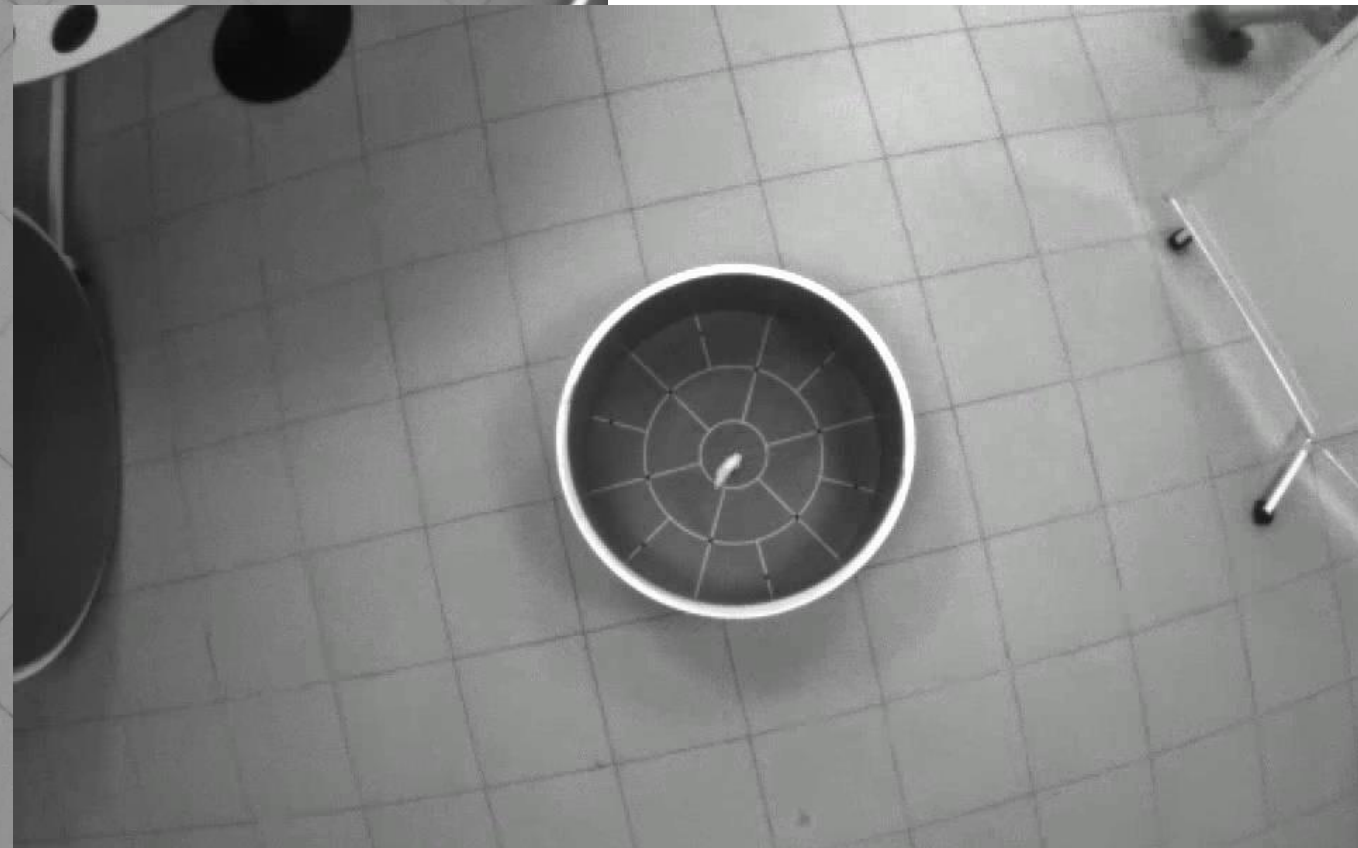


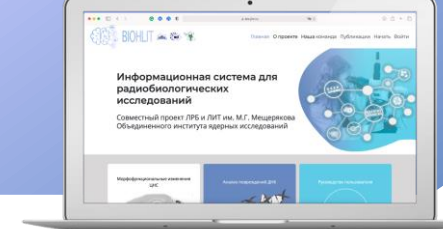
Input data

Videos of different formats and resolutions.

Total number: 36

- (1024, 1280, 3)
- (768, 1024, 3)
- (1080, 1920, 3)
- (960, 1280, 3)





Идея:

Поиск ключевых точек автомобильных номеров с использованием свёрточной нейронной сети



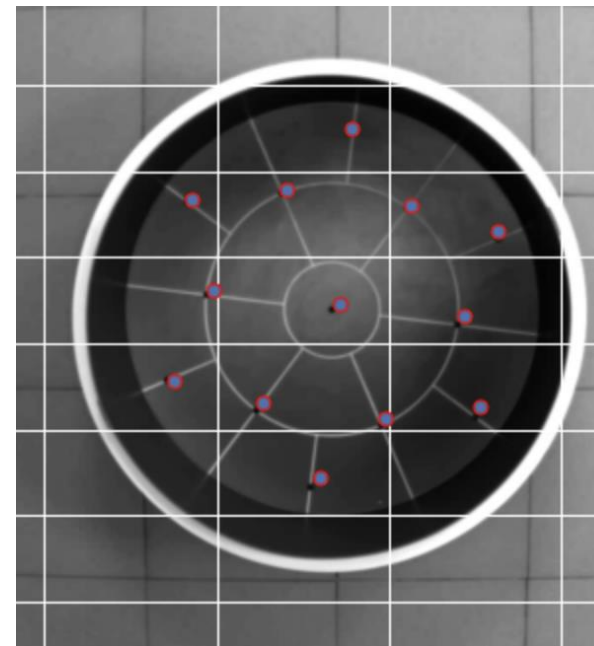
Resource:



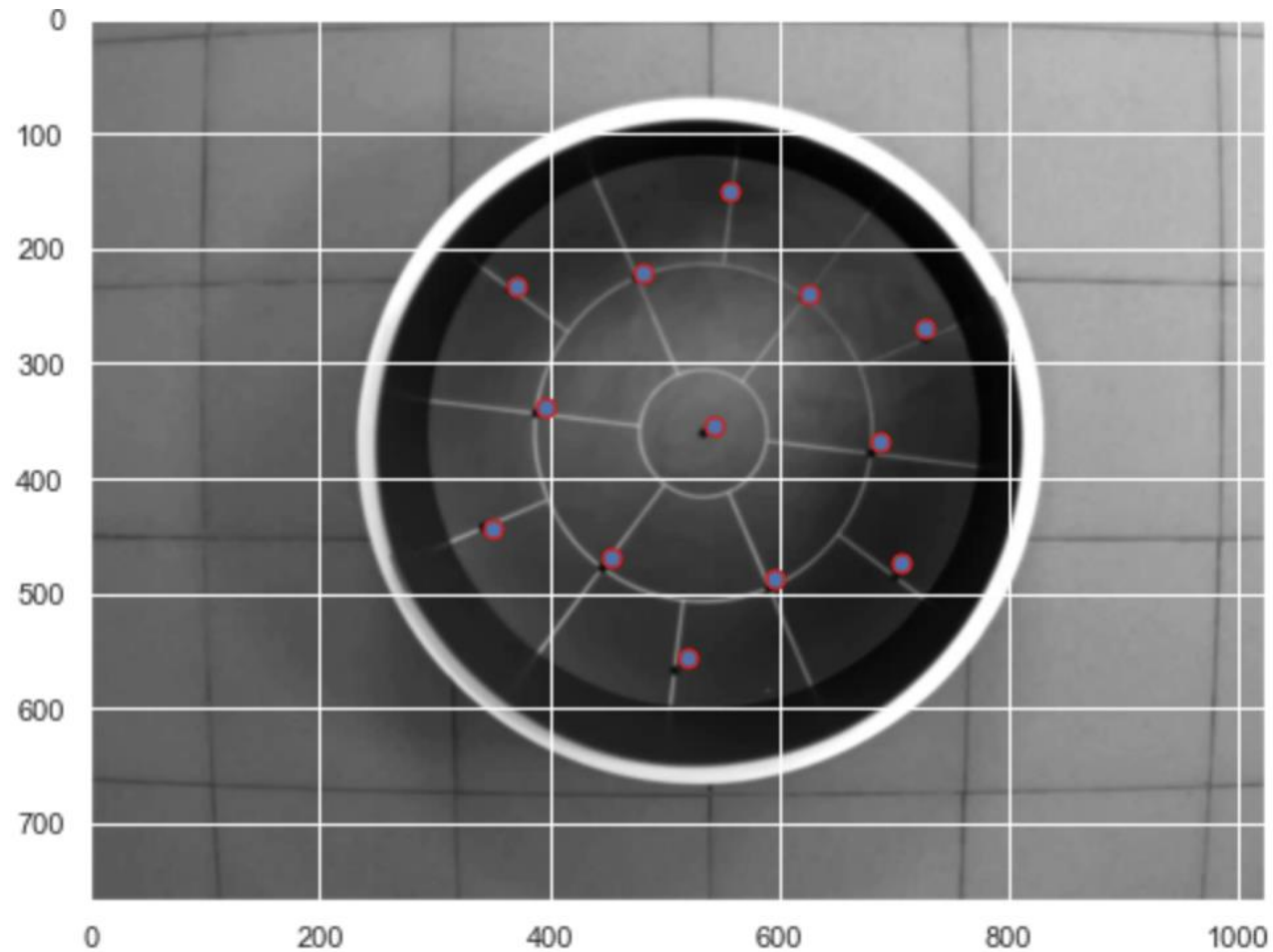
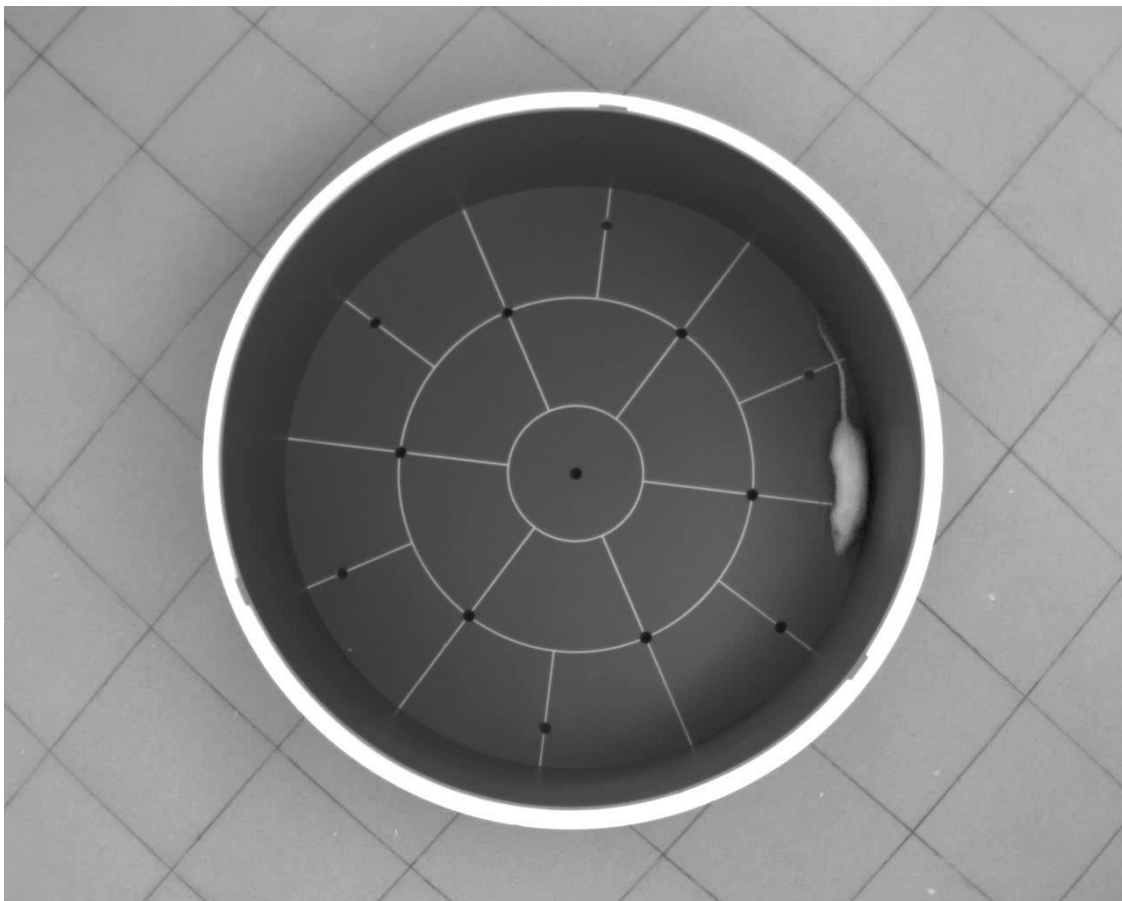
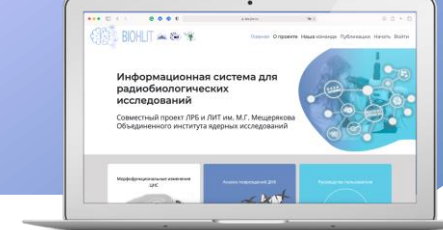
Входные параметры: изображения с ареной;

Выходные параметры: координаты ключевых точек, для которых выбираются отверстия («норки»).

Для обучения сверточной нейронной сети подготовлен размеченный набор данных с координатами характерных точек, расположенных в одинаковом порядке.

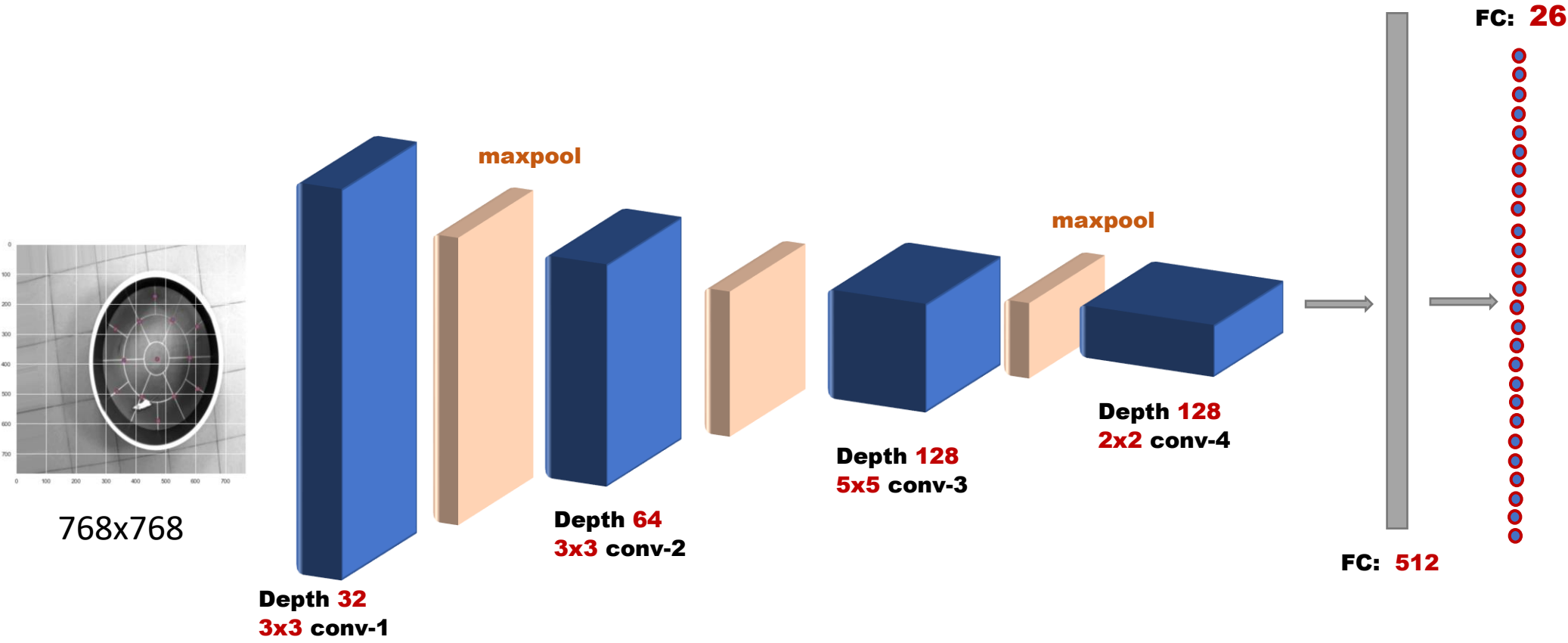


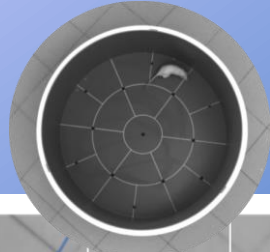
+ CSV



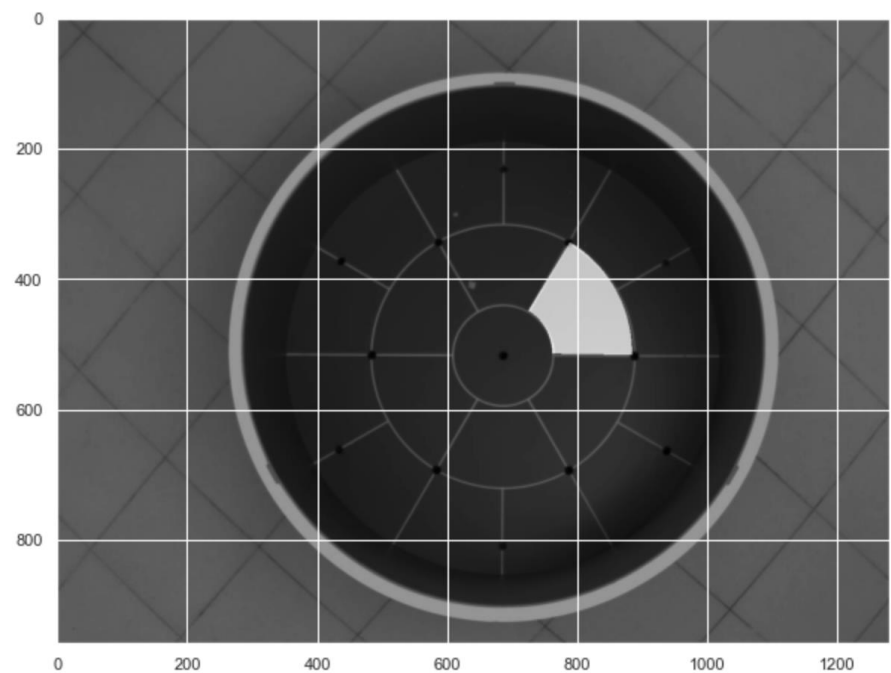
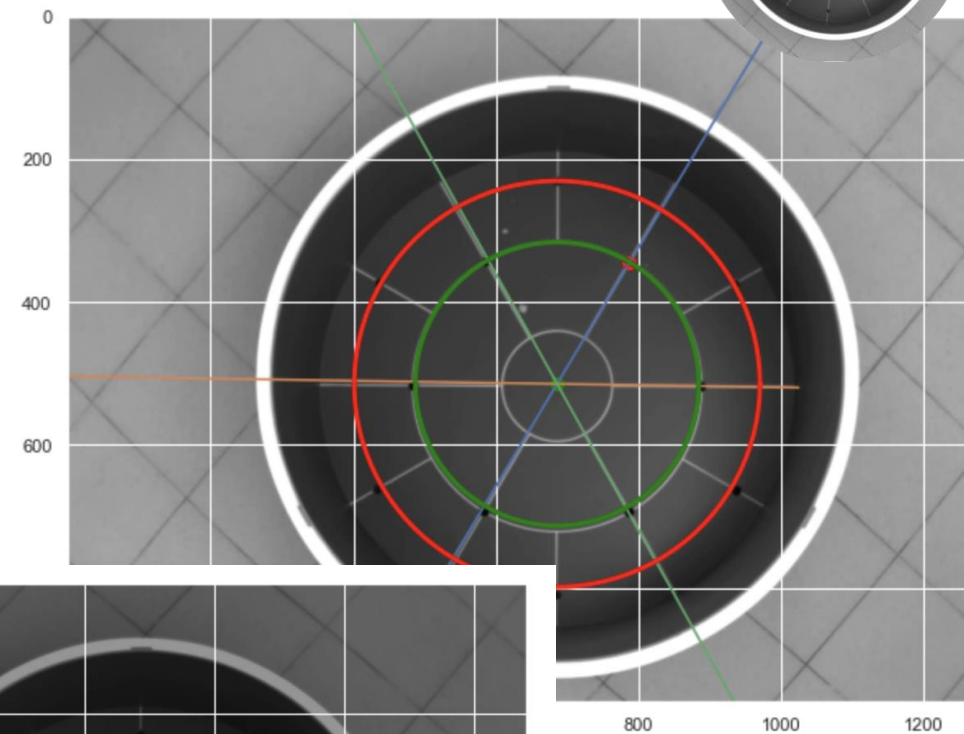
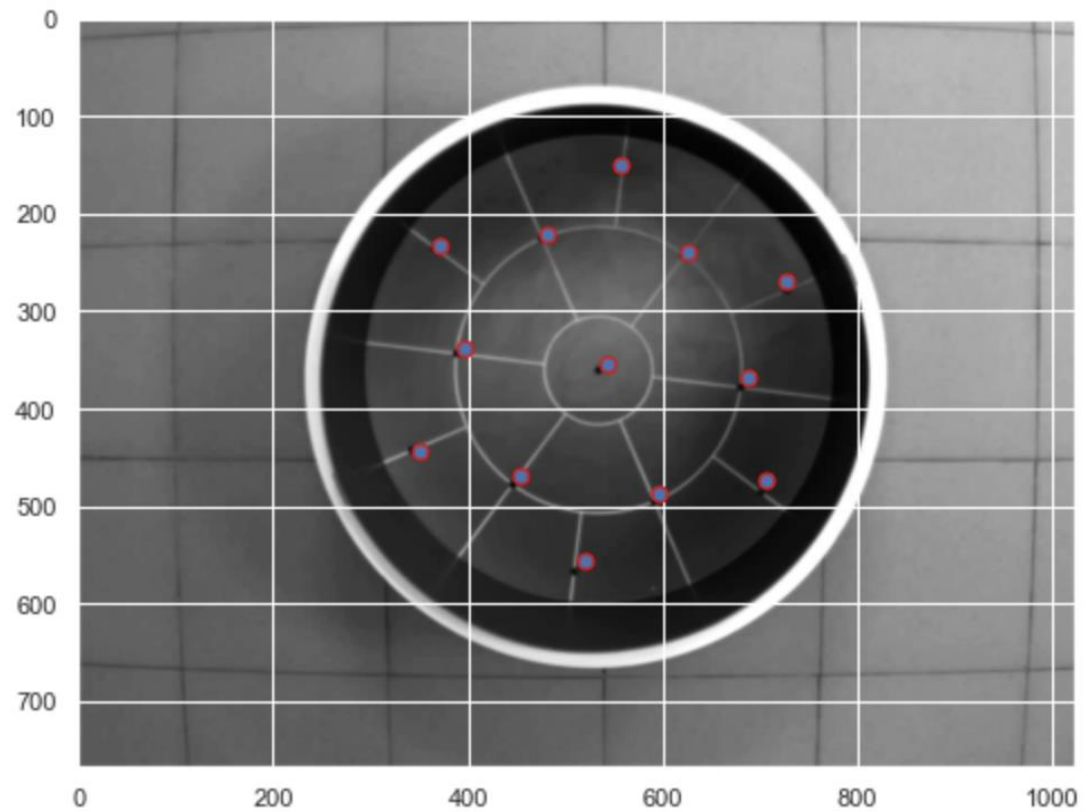


Convolutional Neural network architecture

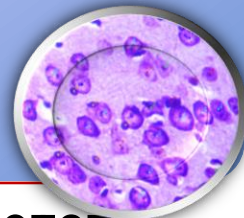




0
(768, 1024)



Stage #2: изучение морфологических изменений в клетках центральной нервной системы

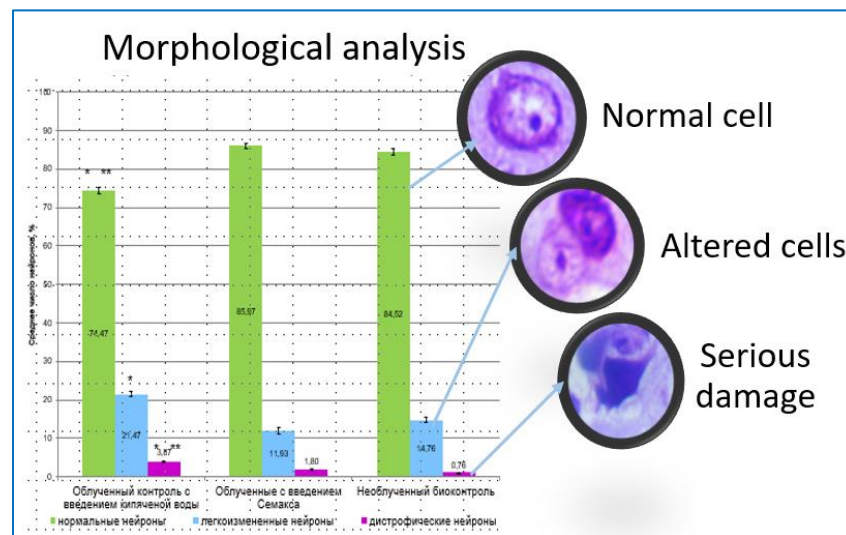
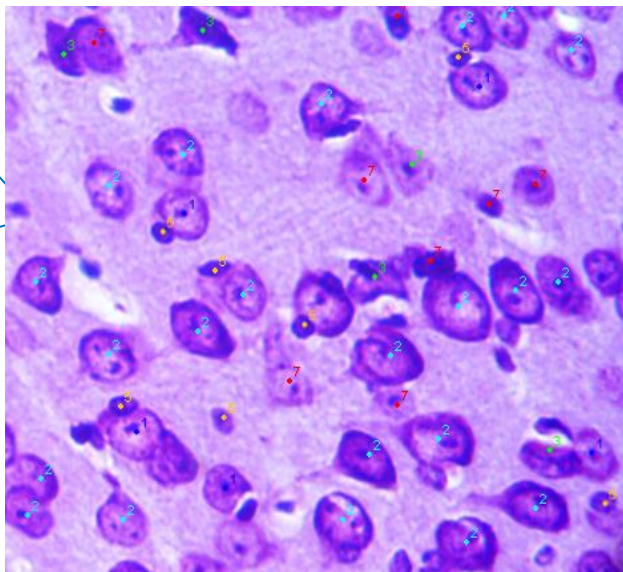


Второй этап направлен на изучение морфологических изменений в клетках центральной нервной системы для установления связи с поведенческими реакциями лабораторных животных, полученными на первом этапе исследований. Проводят также гистологическое исследование других органов лабораторных животных.

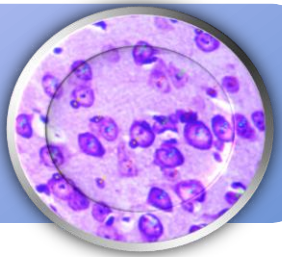


С точки зрения анализа данных этот этап связан с **анализом изображений** срезов нервной ткани разных отделов головного мозга, полученных с помощью светового микроскопа, фотоаппарата и программ, позволяющих подсчитывать клетки с определенными изменениями.

В рамках проекта решается задача автоматизации морфологического анализа гистологических препаратов путем реализации алгоритмов, основанных на нейросетевом подходе и методах компьютерного зрения.



Stage #2: изучение морфологических изменений в клетках центральной нервной системы

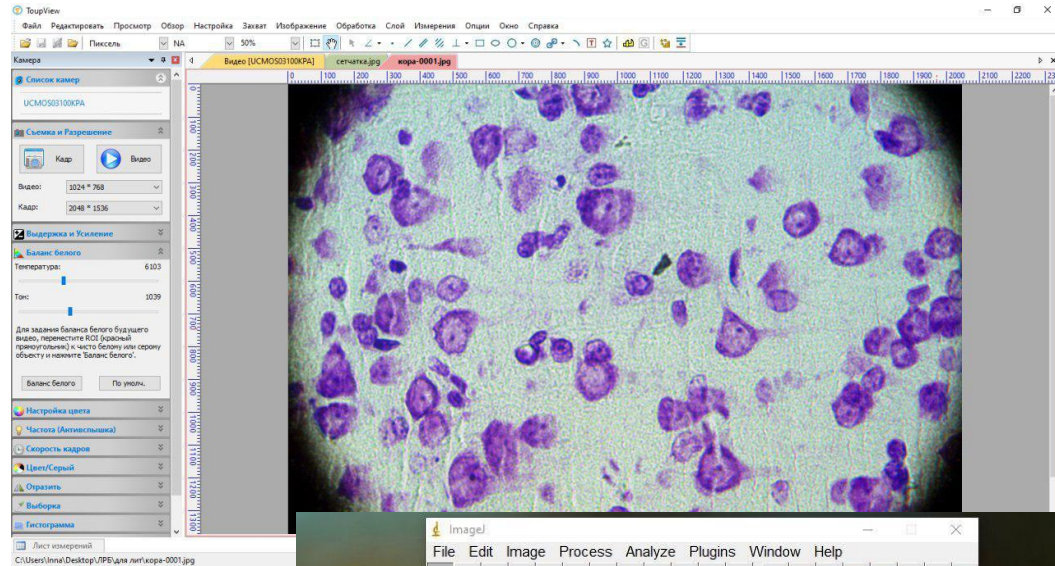


Stages of histological research

Light microscope

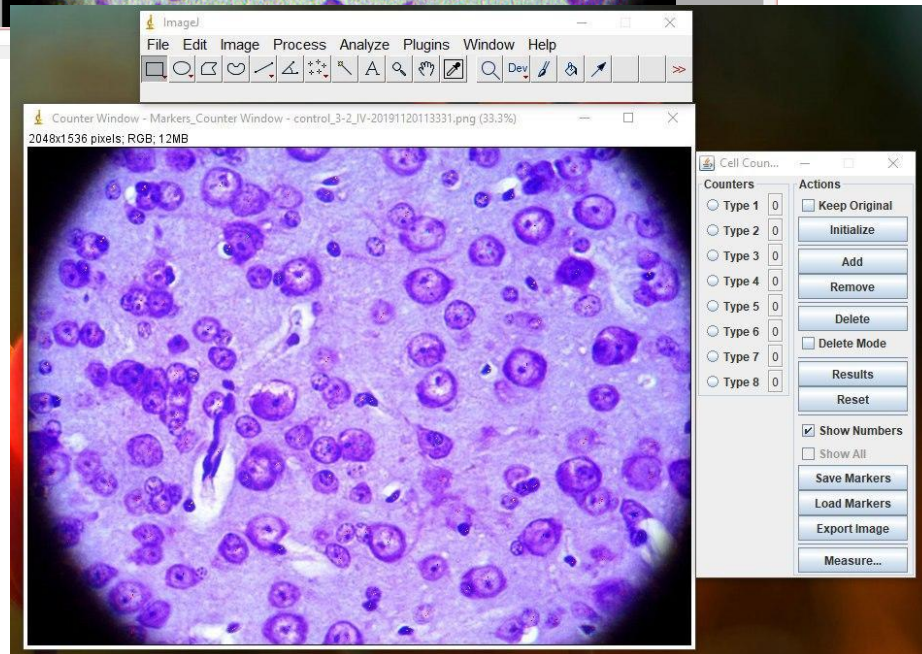


Histological preparations



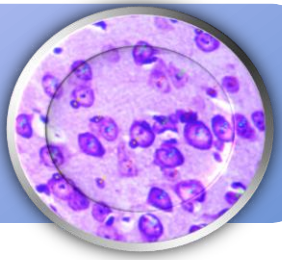
M. 40x10. Histological staining of the nervous tissue by the Nissl method

ToupView



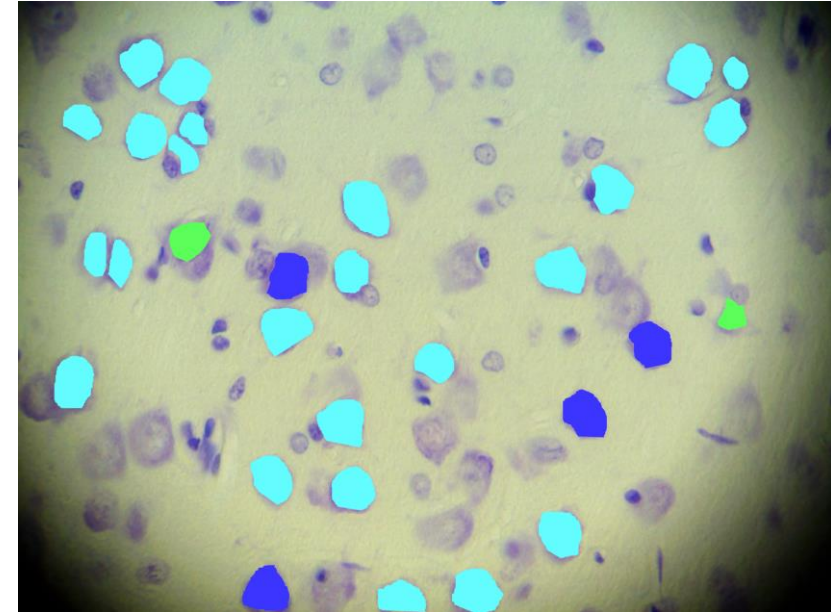
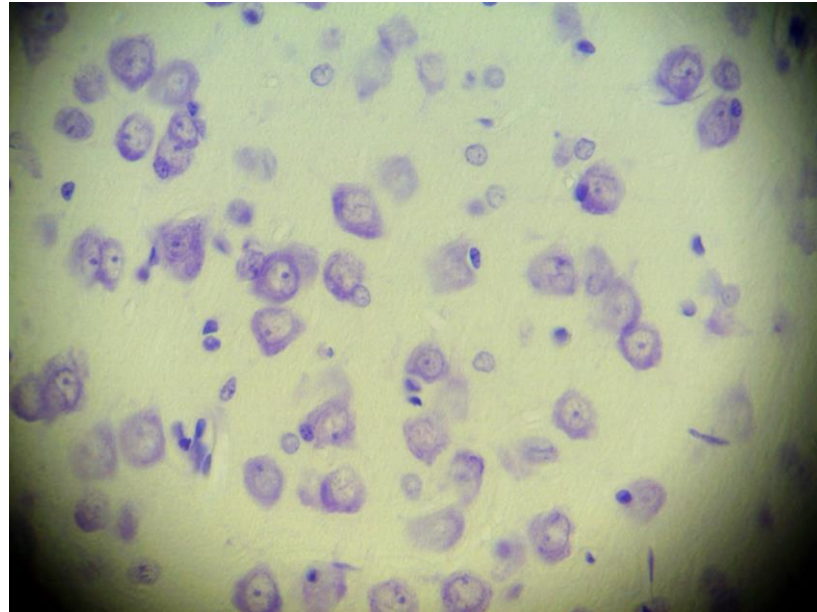
ImageJ

Stage #2: изучение морфологических изменений в клетках центральной нервной системы

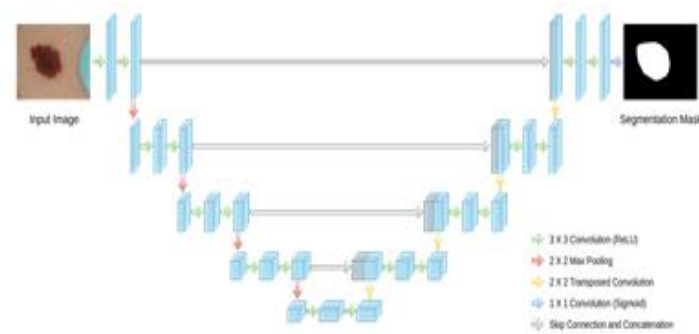


Used classification of brain cells:

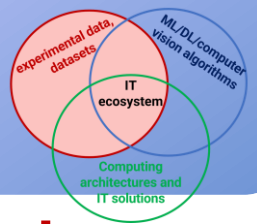
- normal (1);
- easily modified (2): morphofunctional and compensatory-adaptive;
- degenerative (3);
- glia (5);
- elements of the nervous tissue that are difficult to identify (7)



For the segmentation task, the U-net neural network architecture is well suited



IT ecosystem: experimental data, datasets



experimental data

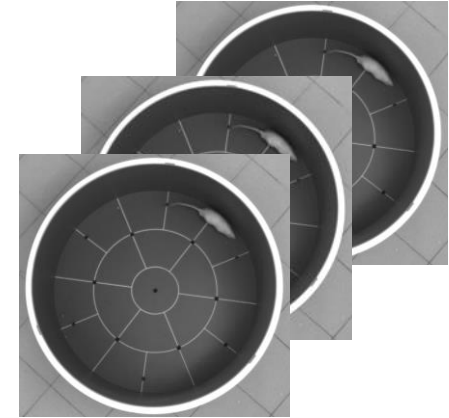
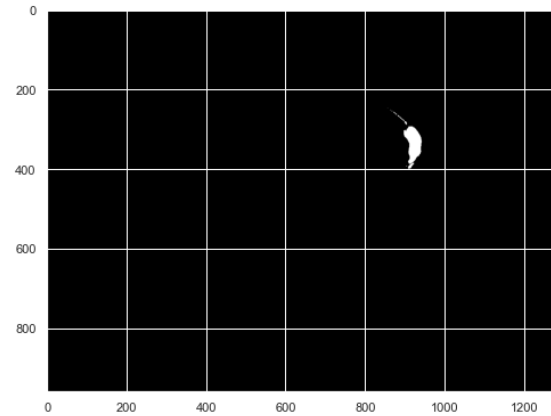


Video files



data labeling/annotation tool

labeled/ annotated data



Action recognition

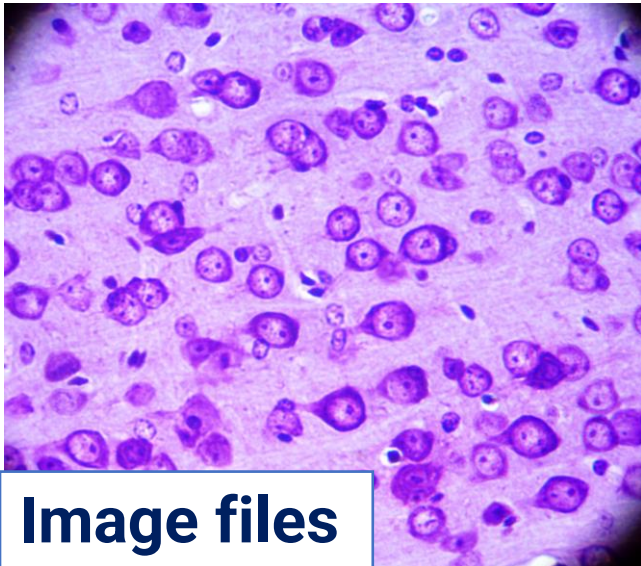
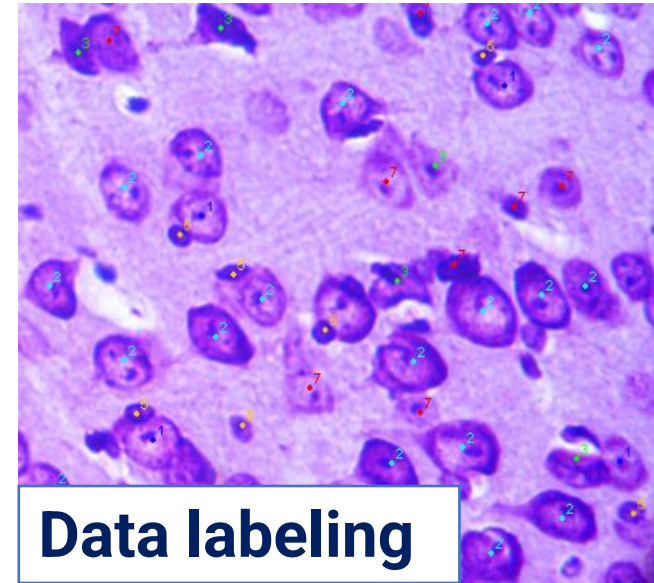


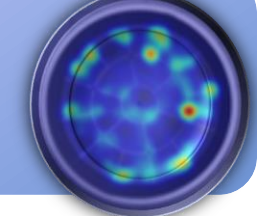
Image files



data labeling/annotation tool



Data labeling



Эксперимент:

[✎ Редактировать](#)

вьетнам

Описание:

лечение

Дата забоя:

28-06-2022

Тип животных:

mouse

Пол животных:

Дата завоза:

28-06-2022

Дата рождения животных

28-06-2022

Дата облучения:

28-06-2022

Облучение

Доза Gr

5Гр

Другие воздействия

Поведение животных

[Открытое поле](#) [Моррис](#) [Т-лабиринт](#)

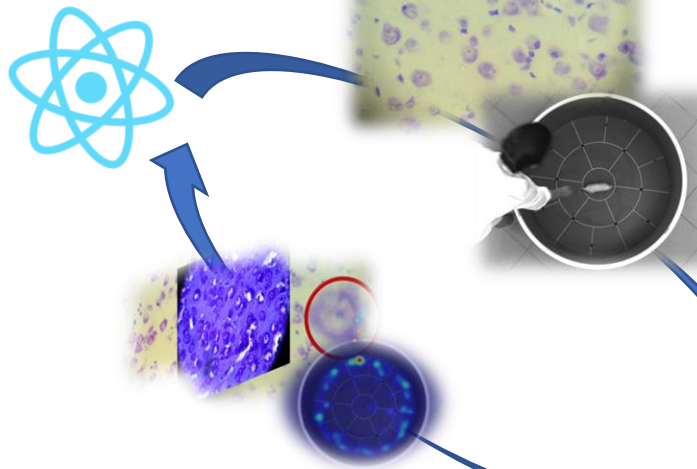
[Загрузить](#)

Всего видео:7

Выбраны:

<input type="checkbox"/> №	Скачать	Имя файла	Статус	Запущен	Завершено	Анализ группы
<input type="checkbox"/> 1	Ссылка	Control_mouse1 10-22-30.avi	Running/Pending	2022-07-06 13:15:15		Анализ Результаты Удалить файл
<input type="checkbox"/> 2	Ссылка	Control_mouse2 10-31-24.avi	FAIL	2022-06-28 11:25:36	2022-06-28 11:25:52	Анализ Результаты Удалить файл
<input type="checkbox"/> 3	Ссылка	Control_mouse3 10-38-39.avi	FAIL	2022-06-28 11:27:16	2022-06-28 11:27:33	Анализ Результаты Удалить файл

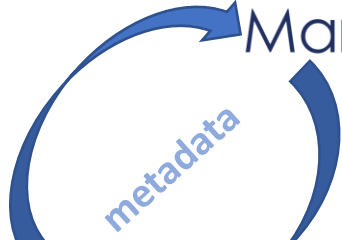
Webapp



Metadata DB



MariaDB®



API-Server

Auth data, credentials



Auth

Supercomputer "Govorun"



Train, Inference, Results

slurm workload manager
Batch-system

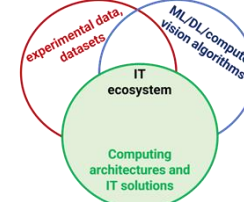
Train, Inference, Dev

jupyterhub

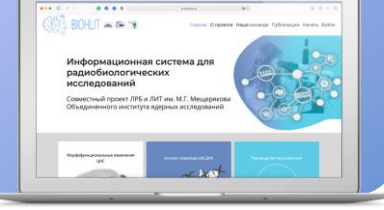
ML\DL ecosystem

HLIT-storage, JINR EOS

Images, Videos, data



IT ecosystem



Computational field

Tiered storage system

Databases

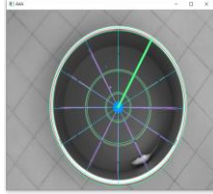
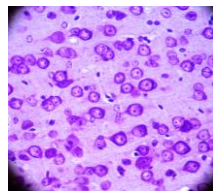
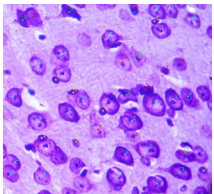
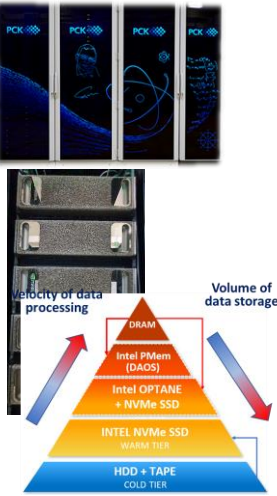
Algorithmic block for behavioral analytics

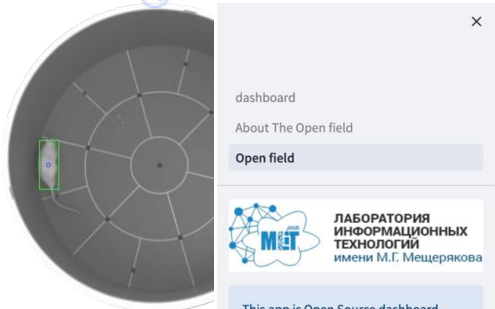
Algorithmic block for histological studies

Algorithmic block for a comprehensive study of morphofunctional changes

Data markup and annotation services

Web services for data loading, data analysis





dashboard

About The Open field

Open field

ЛАБОРАТОРИЯ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ
имени М.Г. Мещерякова

This app is Open Source dashboard.

Site of MLIT JINR: [Link](#).

the project is being created within the framework of the ML/DL/HPC ecosystem of the HybriLIT platform. Link: [here](#).



dashboard

About The Open field

Open field

ЛАБОРАТОРИЯ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ
имени М.Г. Мещерякова

This app is Open Source dashboard.

Site of MLIT JINR: [Link](#).

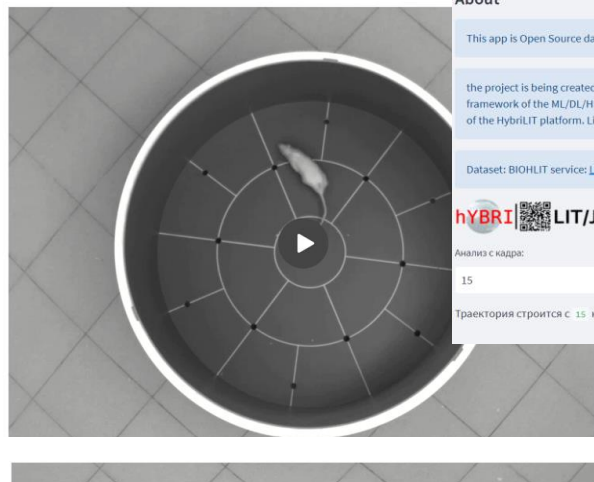
the project is being created within the framework of the ML/DL/HPC ecosystem of the HybriLIT platform. Link: [here](#).



The Open field test-system analysis

Original file

openfield.mp4



```
{
  "FileName": "openfield.mp4"
  "FileType": "video/mp4"
}
```

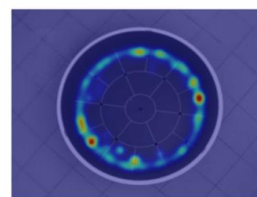
Количество кадров = 500

(500, 1024, 1280)

Analyse

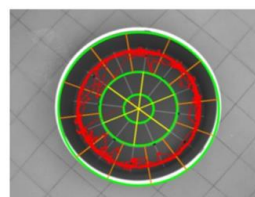
heatmap and trajectory:

Heatmap



Download heatmap

Trajectory



Download trajectory

dashboard

About

Morris water tracking

ЛАБОРАТОРИЯ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ
имени М.Г. Мещерякова

About

This app is Open Source dashboard.

the project is being created within the framework of the ML/DL/HPC ecosystem of the HybriLIT platform. Link: [here](#).

Dataset: BIOHLIT service: [Link](#).

hYBRI LIT/JINR

Анализ с кадра:

15

Траектория строится с 15 кадра

MOUSE TRACK ANALYSIS DASHBOARD

Morris water maze

Upload file

Drag and drop file here
Limit: 200MB per file • MP4, MOV, AVI [Browse files](#)

rat4 12-32-20.avi 1.9MB

752 489 30

имени М.Г. Мещерякова

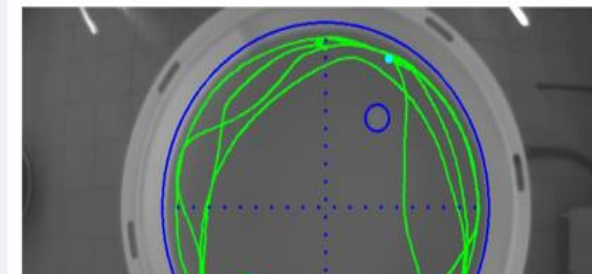
About

This app is Open Source dashboard.

the project is being created within the framework of the ML/DL/HPC ecosystem of the HybriLIT platform. Link: [here](#).

Dataset: BIOHLIT service: [Link](#).

Траектория

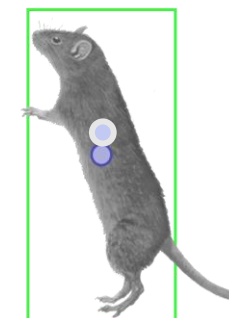
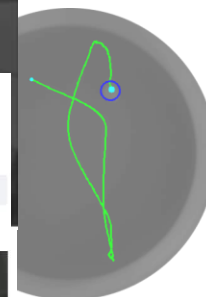
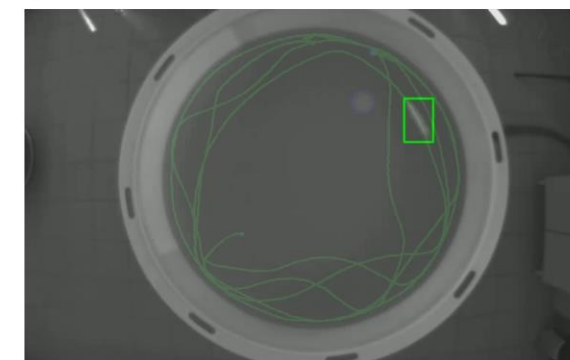


Видео файл для проверки правильности построенной траектории

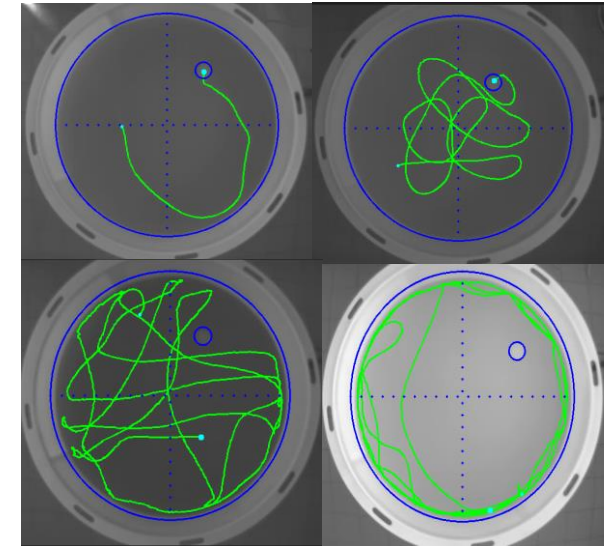
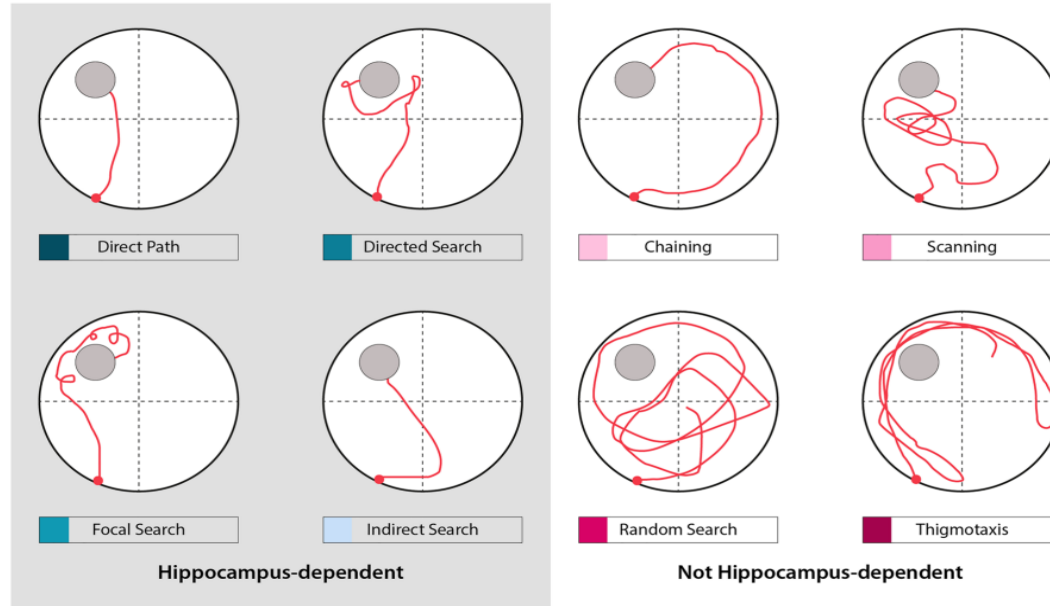
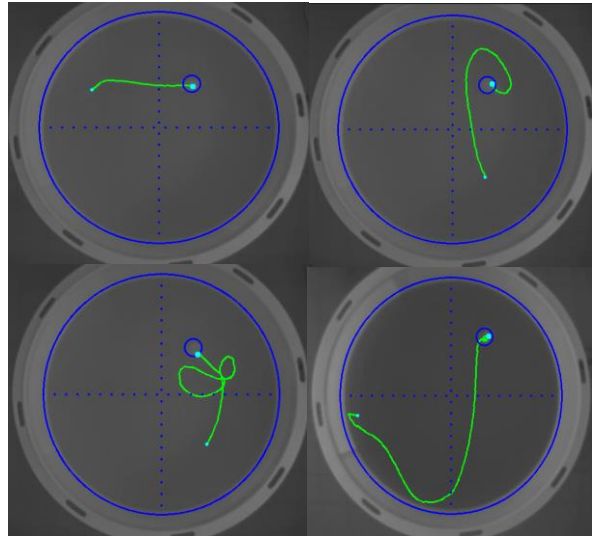
Запись видеофайла с FPS =

30

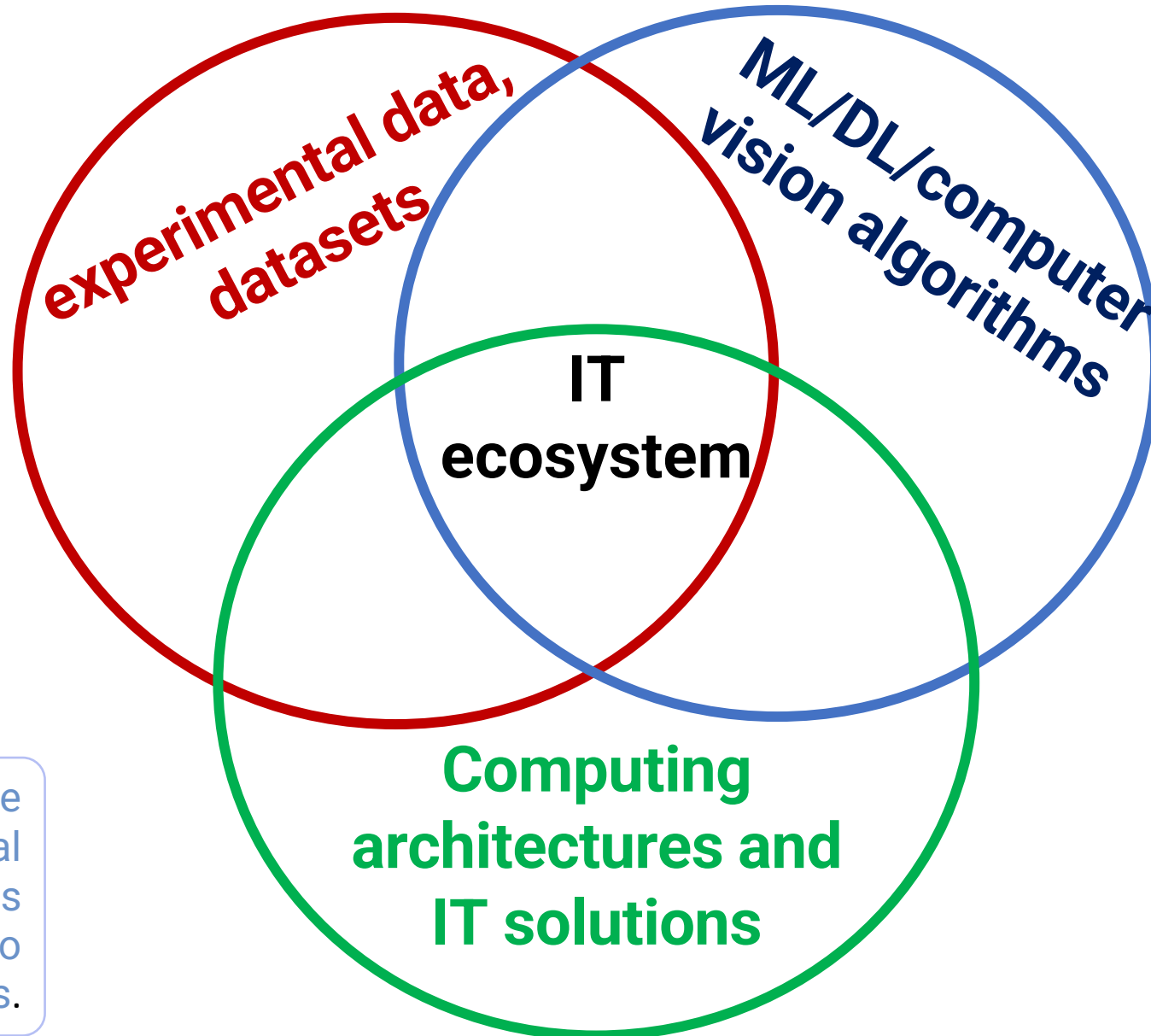
The current FPS is 30



Сервис для аннотирования данных: задача классификации траекторий лабораторных животных в тест-системе «Водный лабиринт Морриса»



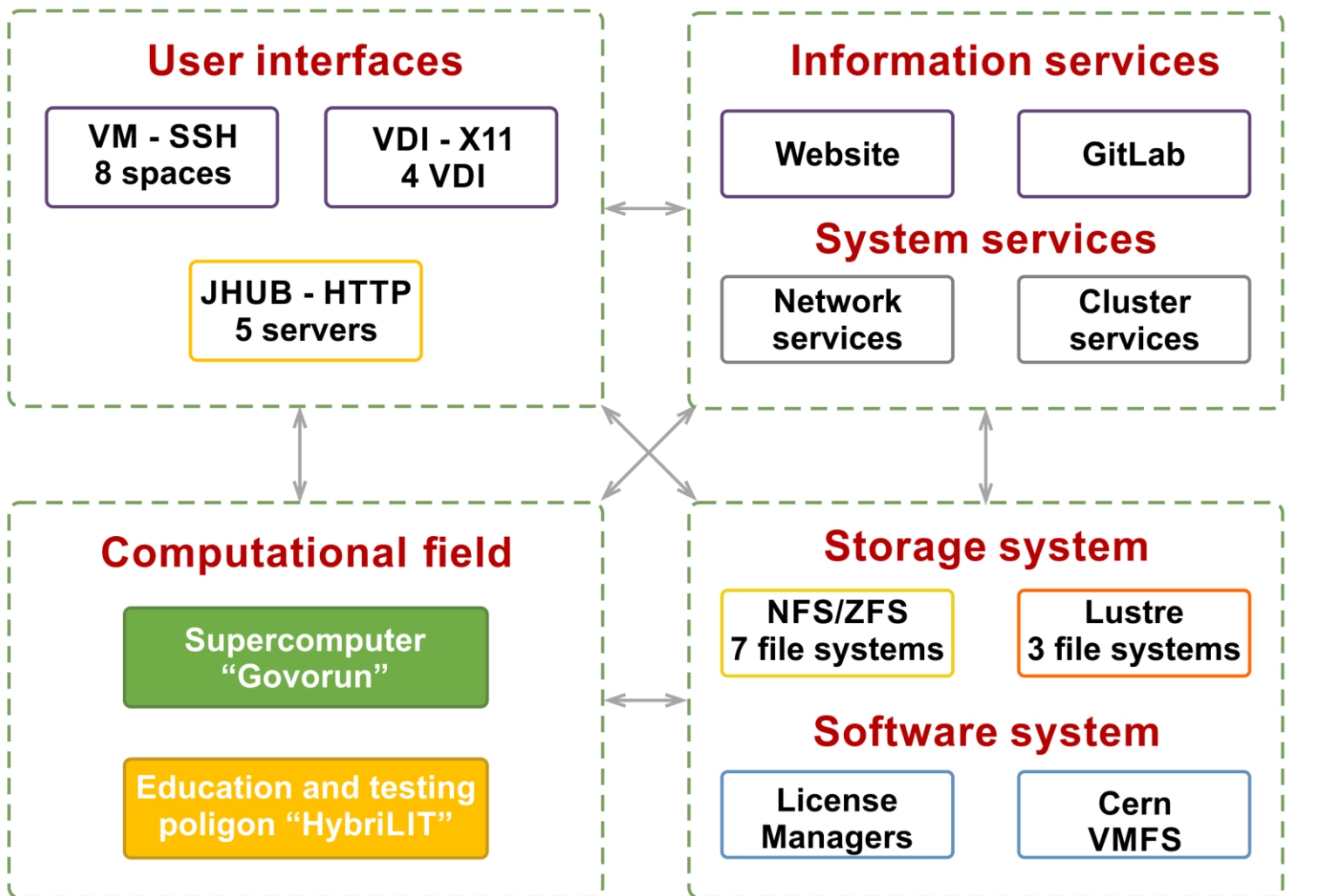
Search strategy analysis of Tg4-42 Alzheimer Mice in the Morris Water Maze reveals early spatial navigation deficits Nadine Curdt, Franziska Schmitt, Caroline Bouter at all Springer Nature Logo March 2022 Scientific Reports 12(1):5451 DOI: 10.1038/s41598-022-09270-1 LicenseCC BY 4.0



Venn diagram on the implementation of the neural network approach, methods and algorithms of ML/DL to solve applied tasks.



Unified software-hardware environment



Экосистема
ML/DL/HPC
платформы
HybriLIT



Полигон для
КВАНТОВЫХ
ВЫЧИСЛЕНИЙ



Cluster HybriLIT 2014:
Full peak performance:
140 TFlops for single precision;
50 TFlops for double precision



#18 в Top50

“Govorun” supercomputer
First stage **2018:**
Full peak performance :
1 PFlops for single precision
500 TFlops for double precision
9th in the current edition of the **IO500**
list (July 2018)



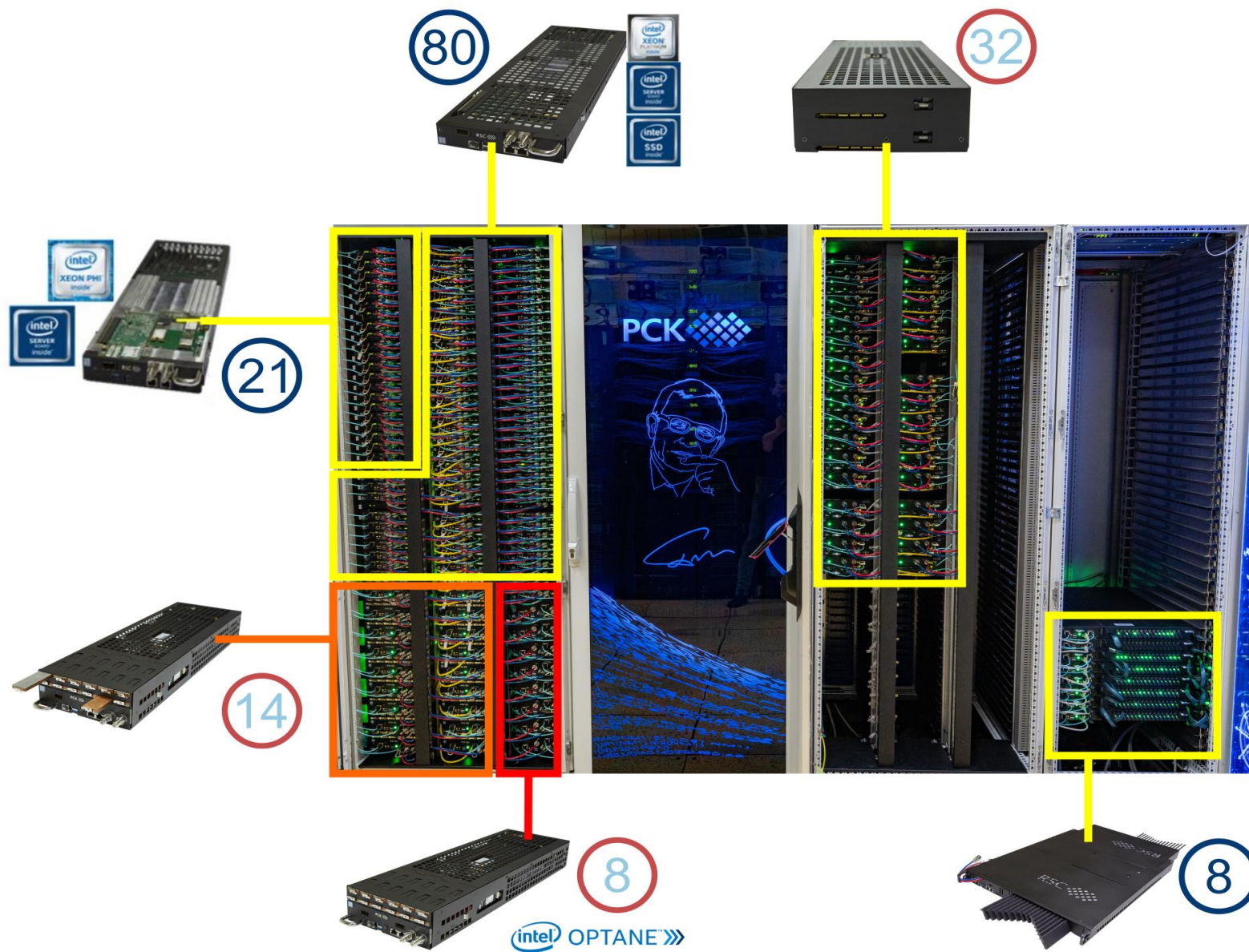
#10 в Top50

“Govorun” supercomputer
Second stage **2019:**
Full peak performance :
1.7 PFlops for single precision
860 TFlops for double precision
288 TB CCXД with I/O speed **>300 Gb/s**
17th in the current edition of the **IO500** list
(July 2020)

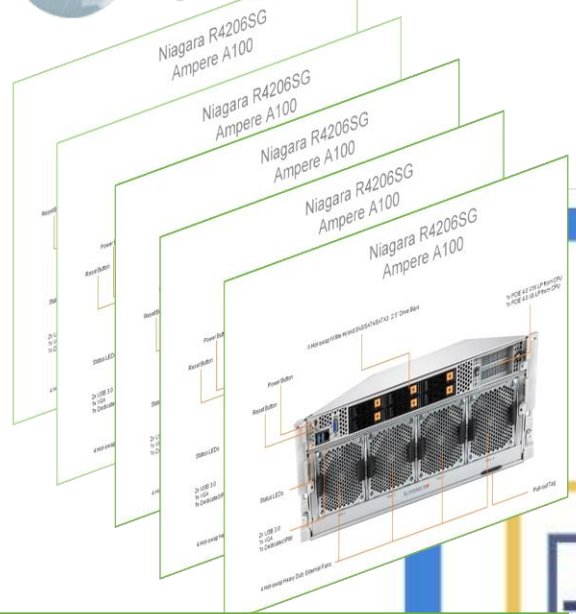


Russian DC Awards 2020 in
“The Best IT Solution for Data
Centers”

Суперкомпьютер «Говорун»: CPU-компонента



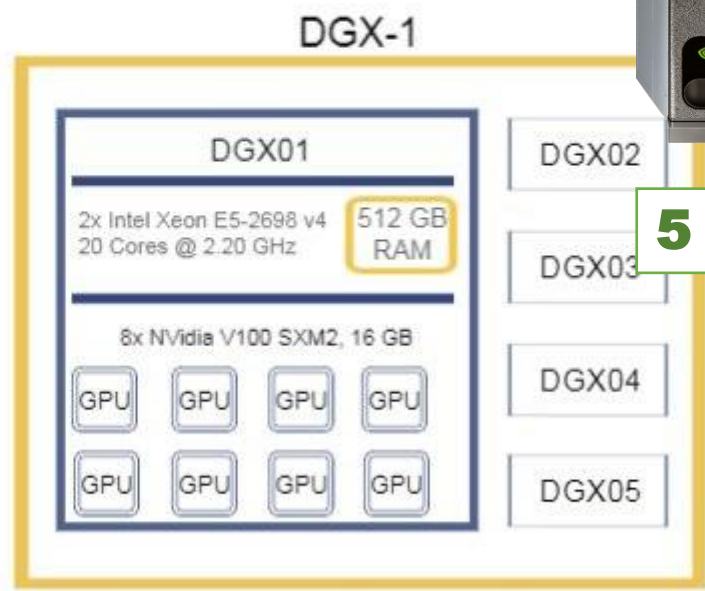
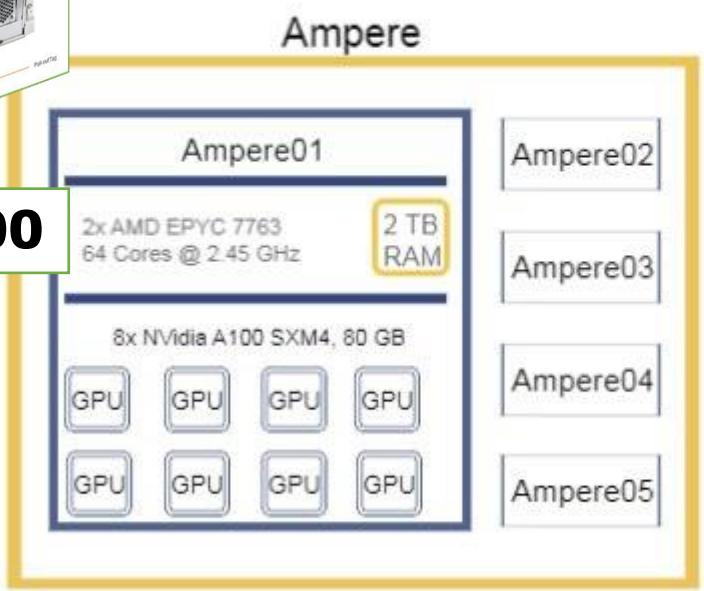
80 GPU NVIDIA



5 x 8 NVIDIA A100

NEW 2023

СК "Говорун" GPU компонента



5 x 8 NVIDIA V100





Warm NFS/ZFS

Students Home
store[1].hydra.local

Backups
store[2].hydra.local

Lustre MLIT-LHEP
store[3-4].hydra.local (MLIT)
lustre[01-02].jinr.ru (LHEP)

User Home (old)
store[5].hydra.local

MPD data storage
store[6].hydra.local

BMN data storage
store[7].hydra.local

Warm NFS/ZFS

User Scratch (ZFS)
s03p002.gvr.local
/zfs/scratch

Warm Lustre

User Home
s03p001.gvr.local
/lustre/home

Hot Lustre

User Scratch (Lustre)
«Lustre 12x12»
s01p[01-02].gvr.local
s02[01-12].gvr.local
/lustre/stor1

MPD data storage
«Lustre ruler x4»
s03p[03-06].gvr.local
/lustre/stor2

Data storage system

8.6 PB

HLIT-VDI (Virtual Desktop Infrastructure)

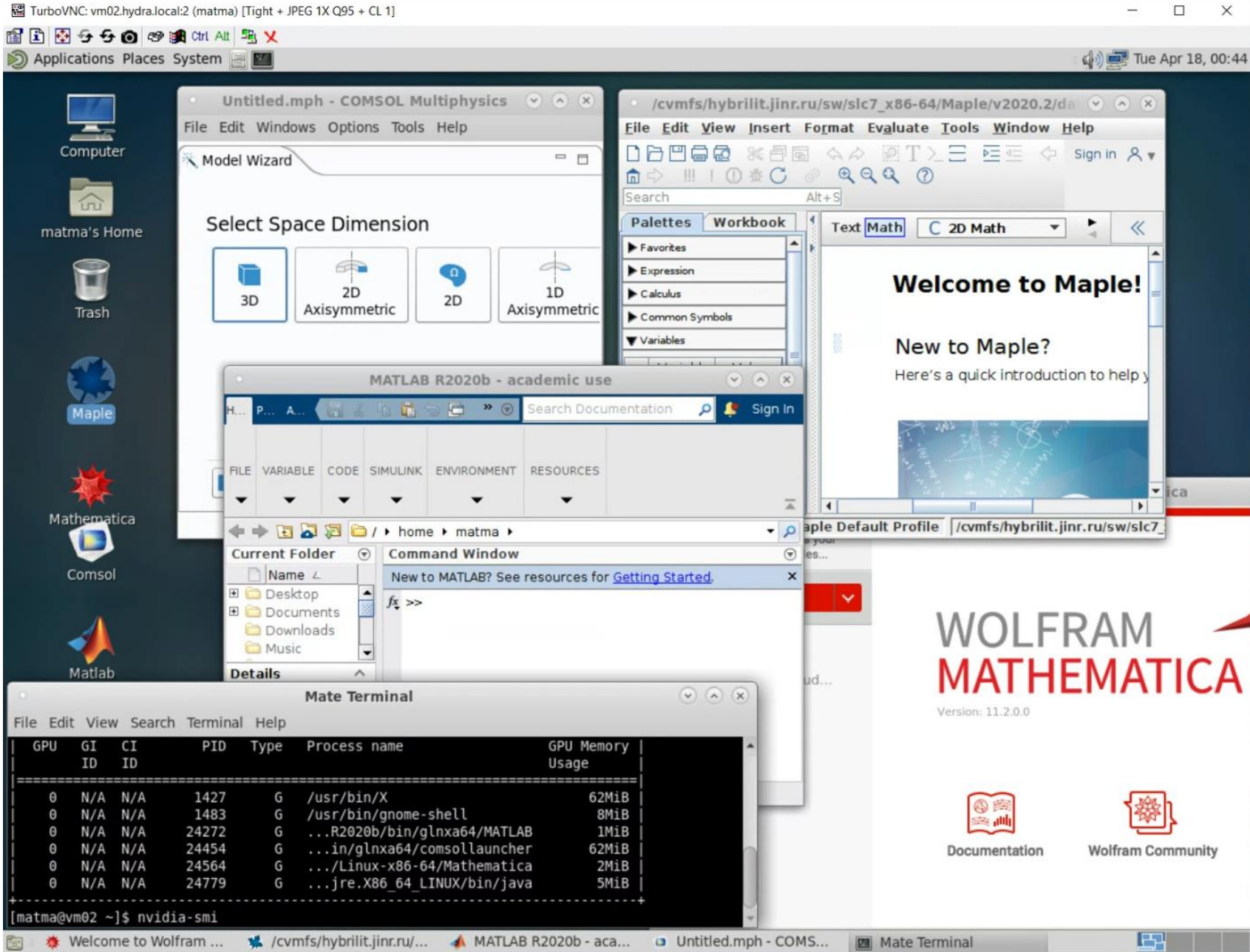
4 VIRTUAL MACHINES

Centos 7.9
RAM 24 Gb
Nvidia Tesla M60, 8 GB
1 Gbit / sec

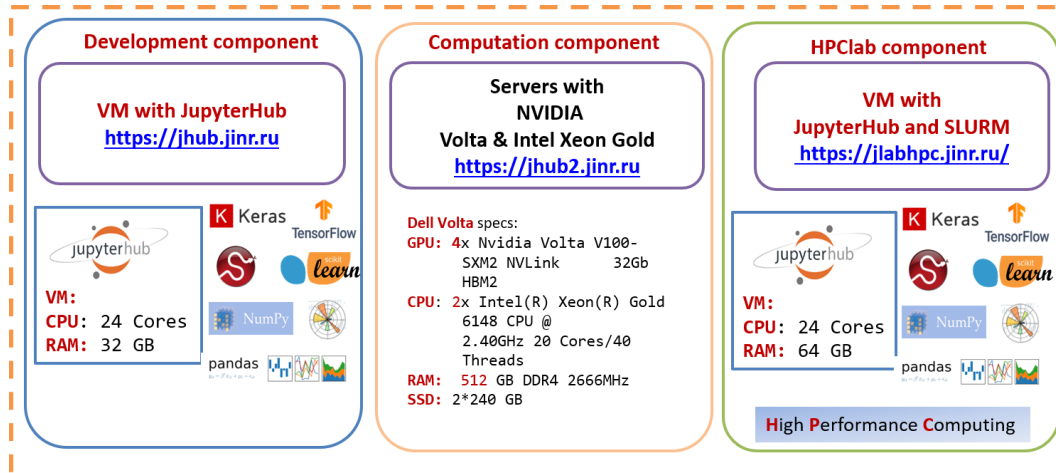
10 USERS

SOFTWARE

Comsol
Maple
Mathematica
Matlab



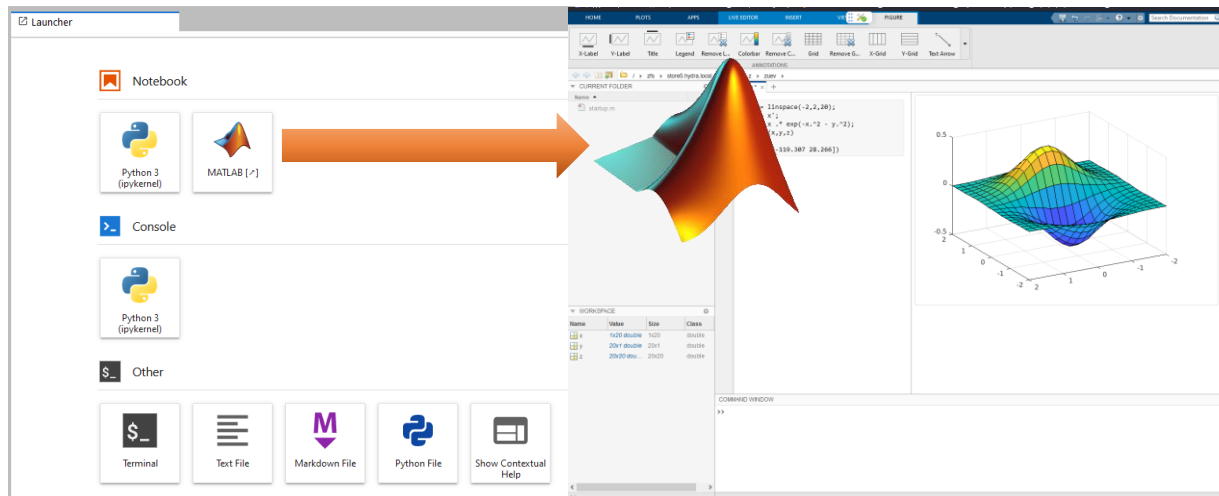
ML/DL/HPC Ecosystem of the HybriLIT Heterogeneous Platform: New Opportunities for Applied Research



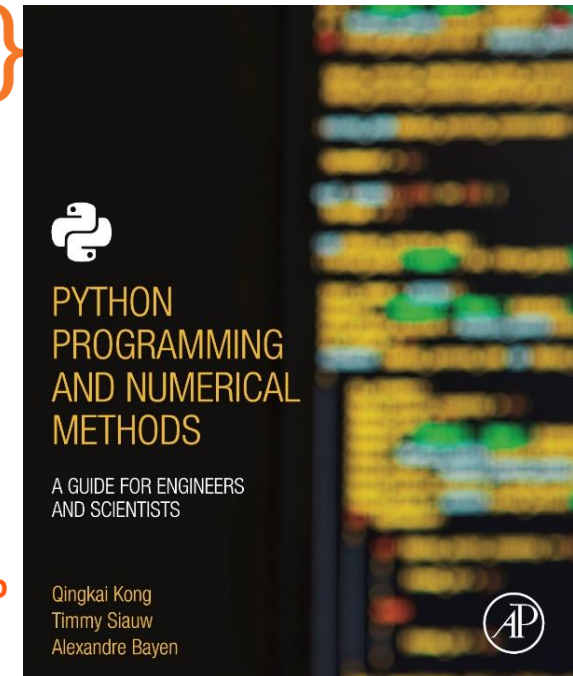
The **ML/DL/HPC ecosystem** is now actively used for machine and deep learning tasks. At the same time, the accumulated tools and libraries can be more widely used for scientific research, including:

- numerical computations;
- parallel computing on CPUs and GPUs;
- visualization of results;
- accompanying them with the necessary formulas and explanations.

In 2022, on the ML/DL/HPC ecosystem, it became possible to run the **MATLAB code in Jupyter Notebook**, which allows one to effectively perform applied and scientific computations.



Python Numerical Methods



HPCLab component
VM with JupyterHub and SLURM [<https://jlabhpc.jinr.ru>]

- ❑ Intel Xeon Gold 6126 (24 Cores @ 2.6 GHz)
- ❑ 32 GB RAM

Educational component
JupyterLab Server [<https://studhub.jinr.ru>]
[<https://studhub2.jinr.ru>]

- ❑ 2x Intel Xeon Gold 6152 (22 Cores @ 2.1 GHz)
- ❑ 512 GB RAM

Computation component
Server with NVIDIA Volta [<https://jhub1.jinr.ru>]
[<https://jhub2.jinr.ru>]

- ❑ 2x Intel Xeon Gold 6148 (20 Cores @ 2.4 GHz)
- ❑ 4x **NVIDIA Tesla V100** SXM2 32 GB HBM2
- ❑ 512 GB RAM

[<https://jhub3.jinr.ru>]

- ❑ 2x Intel Xeon E5 2698v4 (20 Cores @ 2.2 GHz)
- ❑ 8x **NVIDIA Tesla V100** SXM2 16 GB HBM2
- ❑ 512 GB RAM

Notebook

Python 3 (ipykernel) Cirq PennyLane Qiskit

Console

Python 3 (ipykernel) Cirq PennyLane Qiskit

Other

Terminal Text File Markdown File Python File Show Contextual Help

<https://jhub3.jinr.ru>

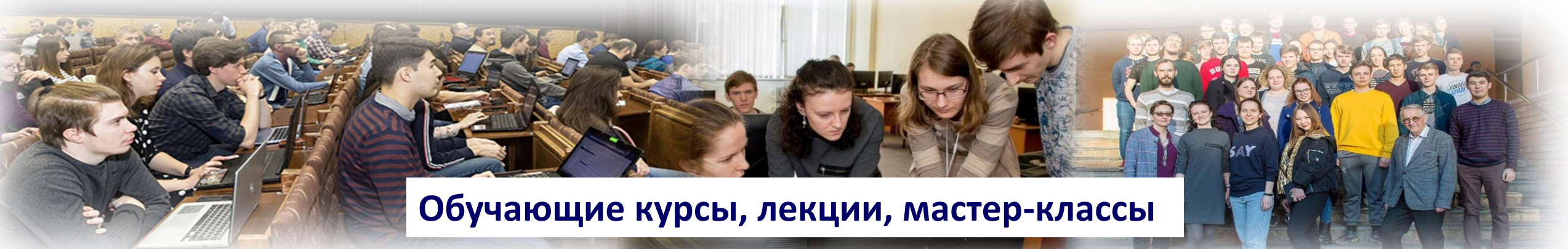
Характеристика сервера:
• 2x Intel Xeon E5 2698v4 (20 Cores @ 2.2 GHz),
• 512 GB RAM,
• 8x NVIDIA Tesla V100 SXM2 16 GB HBM2

```
%matplotlib inline  
circuit.draw(output='mpl')
```

The diagram shows a quantum circuit with four qubits: q0, q1, q2, and C. Qubit q0 starts with an X gate. q1 and q2 have CNOT gates controlled by q0. q1 has an H gate. q2 has a CNOT gate controlled by q1. q0 has an H gate. q1 and q2 have CNOT gates controlled by q0. q2 has a Z gate. All three qubits (q0, q1, q2) are measured at the end. The measurement results are labeled 0, 1, and 2.



Образовательная деятельность



Обучающие курсы, лекции, мастер-классы

Сотрудники ЛИТ, ведущие ученые ОИЯИ и стран-участниц

Ведущие производители современных вычислительных архитектур и программного обеспечения

Технологии параллельного программирования



Инструменты для отладки и профилирования параллельных приложений



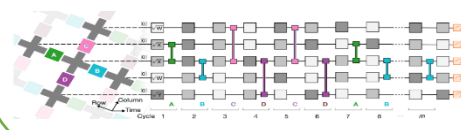
Прикладные пакеты программ



Фреймворки и инструменты для задач ML/DL



Квантовые алгоритмы, программирование и управление





Образовательная деятельность: учебные курсы

Modern information technologies in biology and medicine

The international workshop “Modern information technologies in biology and medicine”

22-24 November 2023.

«Tutorial on the use of Python for tasks in Bio-Medical research»

60 участников



Осенняя Школа по информационным технологиям ОИЯИ 16-20 Октября 2023

«Инструментарий на основе Python-библиотек и экосистемы Jupyter для решения научных и прикладных задач»

Хакатон по параллельным вычислениям
60 участников

210
Участников
в 2023 г.



V Международная летняя школа молодых ученых «Современные информационные технологии для решения научных и прикладных задач» 14-17 Июня 2023.

«Инструментарий на основе Python-библиотек и экосистемы Jupyter для решения научных и прикладных задач”

70 участников



XVI Международная стажировка молодых ученых стран СНГ Май-Июнь 2023

«Как научить компьютер "видеть"»
20 участников



Образовательная деятельность



Кластер HybriLIT обладает развитой программно-информационной средой, что позволяет построить учебные программы на самом современном уровне и дать обучающимся возможность овладеть передовыми ИТ-технологиями.

Для проведения выездных учебных курсов и для проведения семинаров в Университете «Дубна» был развернут «Мобильный кластер HybriLIT»

Мобильный HybriLIT

hydra	node1	node2
Intel Core i9-9900	Intel Core i7-8700T	Intel Core i7-8700T
8 cores	6 cores	6 cores
2 threads - 16 logical core	2 threads - 12 logical core	2 threads - 12 logical core
16 Gb RAM	32 Gb RAM	32 Gb RAM
nvme 2Tb, 500Gb	Nvidia Quadro P1000 4 Gb	Nvidia Quadro P1000 4 Gb
1 Gbit/sec	nvme 500Gb	nvme 500Gb
	1 Gbit/sec	1 Gbit/sec

310
студентов
в 2023 г.

Государственный университет «Дубна» Институт системного анализа и управления

Архитектура и технологии высокопроизводительных систем, **60 студентов**
 Параллельные распределенные вычисления, **90 студентов**
 Языки и технологии анализа данных, **50 студентов**
 Технологии высокопроизводительных вычислений, **30 студентов**

Тверской государственный университет Математический факультет

Программные средства математических вычислений, **20 студентов**

- бакалаврских работ: **5** магистерских диссертаций: **7**

ML/DL/HPC Ecosystem



Осенняя Школа 2023

по информационным технологиям ОИЯИ

16 - 20 Октября



$$\Gamma_{\nu}^{\mu} = O_{\nu}^{\mu} u_{\nu}$$

$$m_{\lambda} = m_0 \exp\left(\frac{q\lambda}{m_0 c^2}\right)$$

$$L = -m_0 c^2 \sqrt{1 - \frac{v^2}{c^2}}$$

$$W_{\mu} = \left(\frac{1}{c} \frac{\partial \lambda}{\partial t}, \nabla \lambda\right)$$

$$W_{\nu\mu} = W_{\nu} u_{\mu} - W_{\mu} u_{\nu}$$

$$O_{\nu} = F_{\nu} - \frac{1}{c^2} W_{\nu\mu}$$

$$L = -m_0 c^2 \sqrt{1 - \frac{v^2}{c^2}}$$

$$W_{\nu} = \frac{1}{c} \frac{\partial \lambda}{\partial t} e_{\nu}$$

Санкт-Петербург

52 студента из 11 университетов РФ

Тверь

Дубна

Москва

Тула

Петропавловск-

Камчатский

Владикавказ

Челябинск

Томск

Владивосток

- Российский университет дружбы народов им. Патриса Лумумбы
- Национальный исследовательский ядерный университет «МИФИ»
- Санкт-Петербургский государственный университет
- Государственный университет «Дубна»
- Тверской государственной университет
- Тульский государственный университет
- Северо-Осетинский государственный университет им. К. Л. Хетагурова
- Южно-Уральский государственный университет (НИУ)
- Томский политехнический университет
- Дальневосточный федеральный университет
- Камчатский государственный университет им. Витуса Беринга



Результаты ИТ-Школы ОИЯИ



4 – Сотрудники ЛИТ

4 Магистерские диссертации
Санкт-Петербургский
государственный университет

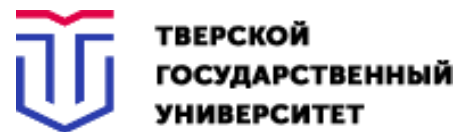
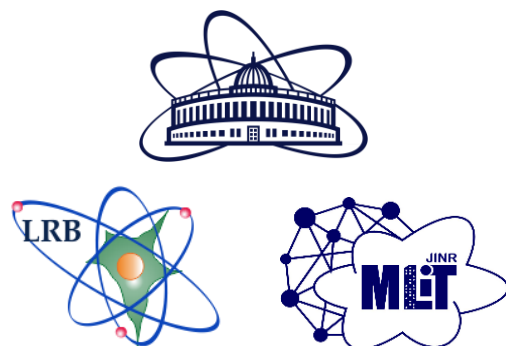
8 Бакалаврские работы: Университет
«Дубна»,
Санкт-Петербургский государственный
университет,
Северо-Осетинский государственный
университет им. К.Л. Хетагурова



Студенческая секция,
в которой приняли участие
16 учащихся из разных вузов России:
МГУ, СПбГУ, РУДН, СОГУ, ТвГУ,
Тульского государственного
университета, Университета «Дубна»

IT SCHOOL
JINR

Весенняя школа по
информационным технологиям ОИЯИ



Спасибо за внимание!