Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение

высшего образования

**«Пермский национальный исследовательский**

**политехнический университет»**

**(ПНИПУ)**

**ОТЧЕТ**

**о прохождении стажировки**

Игошев Матвей Иванович

(фамилия, имя, отчество)

Интеллектуальный анализ данных и технологии машинного обучения для цифровой трансформации деятельности компании

(наименование программы ДПО)

**Проверил:**

Руководитель ДПП \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(оценка) (подпись) (инициалы, фамилия)

**Пермь 2024**

**ВВЕДЕНИЕ**

1. Краткая характеристика места стажировки:

Стажировка проводилась в компании «Датабриз». Компания Датабриз является разработчиком и интегратором приложений и информационно-аналитических систем для корпоративных клиентов и региональных органов власти РФ и стран СНГ.

2. Цели стажировки:

- Разметить данные в заготовленном ранее наборе данных.

- Изучить функционал label studio.

- Получить практические навыки в работе с разметкой данных в label studio.

**ОСНОВНАЯ ЧАСТЬ**

1. Выполнение программы стажировки:

Таблица 1.

|  |  |  |
| --- | --- | --- |
| Запланированные мероприяния | Сроки | Отчет о выполнении |
| Получение задания на практику | 20.03.2024 — 21.03.2024 | Результат представлен в данном отчете |
| Ознакомление с технологиями поставленной задачи | 20.03.2024 — 21.03.2024 | Результат представлен в данном отчете |
| Выполнение основной задачи | 21.03.2024 — 8.04.2024 | Результат представлен в данном отчете |

2. Основные научные, учебно-методические и производственные результаты:

Для того чтобы выполнить разметку, было необходимо изучить функционал Label studio. Label studio — это гибкая платформа для разметки данных для последующего обучения или валидации моделей ИИ.

Label studio позволяет выполнять разметку различными способами. В рамках поставленной задачи было необходимо выполнить разметку областью. Разметка областью подразумевает выделение контура объекта. Притом данный набор данных предполагает выполнение обнаружения объектов, то есть на одном изображении может находиться несколько классов или/и несколько объектов одного класса. Каждой области соответствует свое наименование, которое также иногда называют лейбл(label). Выполнение разметки происходит следующим образом: выбирается класс, далее выполняется расстановка точек, из которых складывается контур. Контуру присвается соответствующий класс.

Сам набор данных представляет собой набор изображений для РЖД. Соответственно содержит в себе следующие классы: башмак,(bashmak) СИЗ(orange\_vest), коробка с песком(sand\_box), вентиль(valve), пирамида(piramida), колесо(koleso)(В скобках указаны оригинальные лейблы, использованные для разметки).

Примеры тех классов, которые нужно было разметить на картинках набора данных:



Рисунок 1 — колесо



Рисунок 2 — башмак



Рисунок 3 — пирамида



Рисунок 4 — вентиль



Рисунок 5 — коробка с песком



Рисунок 6 — СИЗ

Важно отметить, что при разметке класса пирамида нужно было разметить сначала башмаки, которые стоят в ней, потом пирамиду.

Если несколько одинаковых объектов находились рядом, нужно было выделить каждый из них отдельным контуром.

В рамках работы были обработаны картинки со следующими номерами(В сумме было обработано 600 изображений):

3728

3730

3732

3733

3735

3736

3737

3740

3742

3743

3744

3746

3747

3748

3751

3768

3786

3790

3796

3802

3807

3810

3811

3818

3794

3824

3823

3835

3872

4068

4413

4408

4409

4410

6464

6466

6467

6468

6469

6515

6517

6524

6527

6528

6547

6548

6619

6620

6621

6622

6623

6624

6625

6626

6627

6628

6629

6630

6631

6632

6633

6634

6635

6636

6637

6638

6639

6640

6641

6642

6643

6644

6645

6646

6647

6648

6649

6651

6653

6654

6655

6656

6657

6658

6659

6660

6661

6662

6663

6664

6665

6666

6667

6668

6670

6671

6672

6673

6674

6675

6676

6701

6757

6755

6708

6732

6733

6743

6742

6746

6751

6761

6763

6764

6767

6773

6774

6775

6776

6778

6772

6771

6779

6780

6781

6785

6787

6789

6790

6791

6792

6795

6798

6815

6816

6827

6841

6843

6844

6851

6854

6860

6862

6864

6863

6865

6867

6869

6870

6875

6877

6904

6902

6903

6906

6907

6909

6910

6911

6913

6914

6916

6917

6918

6920

6921

6922

6923

6924

6926

6927

6928

6929

6930

6931

6932

6933

6935

6936

6937

6938

6939

6941

6942

6944

6945

6946

6948

6949

6952

6953

6954

6955

6956

6957

6959

6960

6961

6962

6963

6966

6968

6969

6970

6971

6972

6974

6975

6976

6978

6979

6980

6981

6982

6983

6984

6991

6992

6994

6995

6996

6997

6998

6999

7000

7001

7002

7003

7004

7005

7006

7007

7008

7009

7010

7011

7012

7013

7014

7016

7017

7019

7020

7022

7023

7024

7025

7026

7030

7032

7033

7036

7038

7039

7040

7041

7042

7046

7047

7051

7052

7053

7056

7057

7058

7059

7060

7062

7065

7066

7067

7068

7070

7071

7072

7074

7076

7077

7078

7079

7080

7081

7082

7083

7084

7086

7087

7089

7092

7093

7096

7097

7098

7099

7100

7102

7103

7104

7106

7107

7112

9502

9525

9526

9527

9528

9531

9534

9539

9540

9544

9546

9549

9552

9557

9561

9563

9564

9565

9584

9588

9590

9591

9596

9598

9602

9603

9604

9605

9606

9611

9614

9615

9617

9619

9621

9626

9628

9629

9631

9632

9633

9634

9638

9639

9640

9641

9642

9643

9644

9645

9646

9647

9650

9651

9652

9653

9654

9655

9659

9660

9661

9662

9664

9666

9667

9668

9669

9671

9672

9673

9674

9675

9679

9680

9862

9683

9684

9685

9686

9687

9688

9689

9690

9691

9692

9694

9696

9697

9698

9699

9700

9701

9702

9703

9704

9706

9707

9708

9709

9711

9712

9713

9714

9716

9717

9720

9721

9722

9723

9752

9754

9755

9756

9758

9759

9760

9761

9762

9763

9764

9765

9766

9767

9768

9769

9770

9771

9772

9773

9774

9775

9776

9777

9778

9779

9780

9782

9783

9784

9786

9787

9789

9790

9791

9792

9793

9794

9796

9798

9800

9802

9803

9804

9805

9807

9808

9809

9811

9813

9814

9816

9817

9819

9820

9821

9822

9823

9824

9825

9826

9829

9831

9832

9835

9836

9837

9839

9840

9841

9842

9843

9844

9845

9846

9847

9848

9850

9851

9852

9853

9854

9855

9856

9857

9858

9859

9861

9862

9863

9864

9866

9868

9869

9870

9871

9874

9875

9876

9877

9878

9879

9881

9882

9884

9885

9887

9888

9889

9891

9892

9893

9894

9895

9896

9897

9899

9900

9902

9903

9904

9906

9908

9909

9910

9911

9912

9913

9915

9916

9917

9918

9919

9922

9923

9924

9925

9926

9927

9928

9929

9930

9933

9934

9935

9936

9937

9938

9939

9940

9942

9944

9945

9946

9948

9949

9950

9951

9952

9954

9958

9959

9962

9967

9972

9973

9974

9975

9977

9979

9980

9981

9982

9983

9984

9986

9987

9988

9990

9991

9992

9993

9994

9995

9996

9998

9999

10000

10002

10003

10004

**ЗАКЛЮЧЕНИЕ**

1. Краткие выводы по стажировке:

В рамках практики были получены навыки по разметке набора данных для решений компьютерного зрения. Была выполнена разметка 600 изображений на 6 классов. Получен опыт работы с Label studio.

Стажер \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*(подпись стажера)**(расшифровка подписи)*

Дата