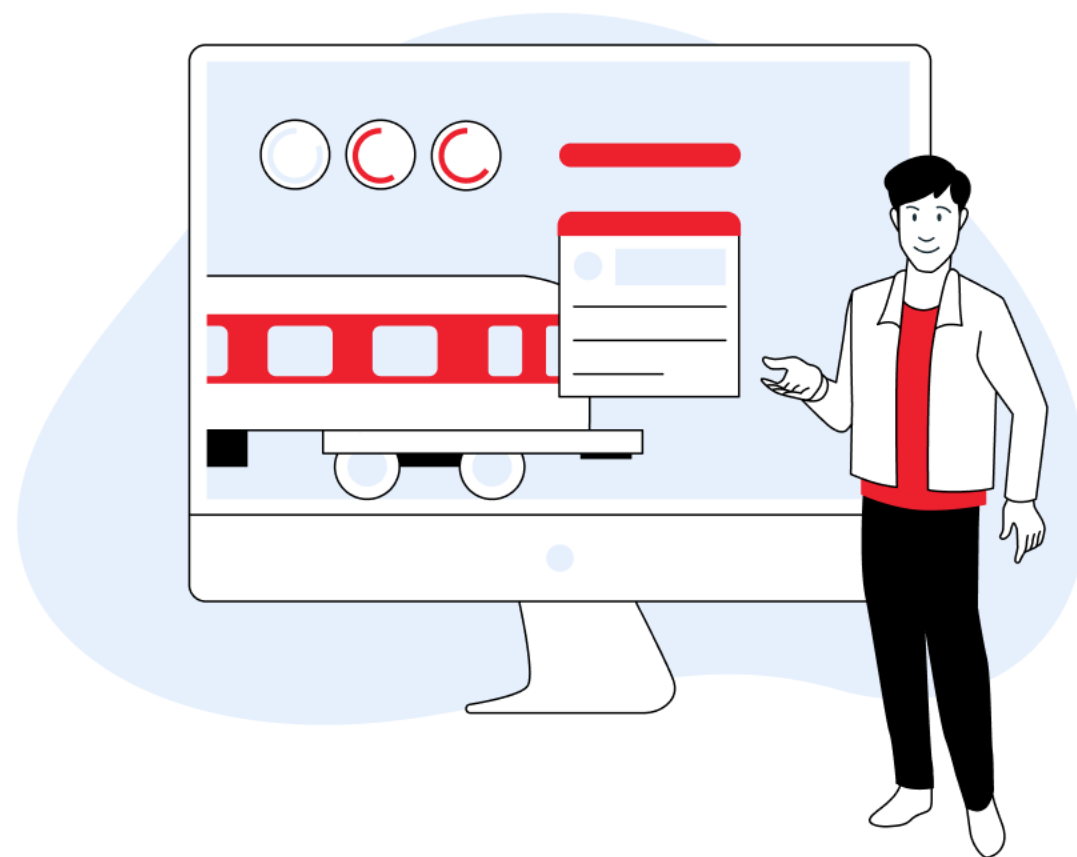


Effectiveness study of machine learning for the task of determining hands position by set of keypoints

Штехин Сергей Евгеньевич

Head of the Computer Vision Group of
AI&NN Laboratory
"OCRv" Sirius dept.

ОЦРВ



This talk: study of determining position of the hands by key points



Using an open dataset using machine learning methods.



Main focus to the development of key features



High quality and compact machine learning model



The effectiveness of various machine learning models were reviewed

Obtained results may be useful in the study labor processes with fast movements and short time intervals for for recognizing technological operations.



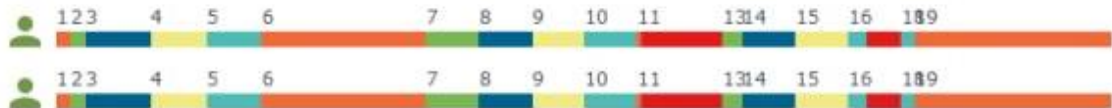
Photo of the working day - site

05

4 Поручение №18. Замер №1



00:00 / 22:15



1. Очистка креплений от грязи

00:00:00 - 00:00:18 (18 с)

1

2

1 Конец шпалы

2. Раскладка резиновых (изолирующих) прокладок по концам шпал

00:00:18 - 00:00:37 (19 с)

1

2

1 Прокладка

3. Отвинчивание и снятие гайки закладного болта

00:00:37 - 00:01:59 (1 м 22 с)

2

2

2 Гайка

4. Снятие двухвитковой и плоской шайб и изолирующей втулки

Technological operations with supplemental tools



Human detection
(Employees)



Tool detection



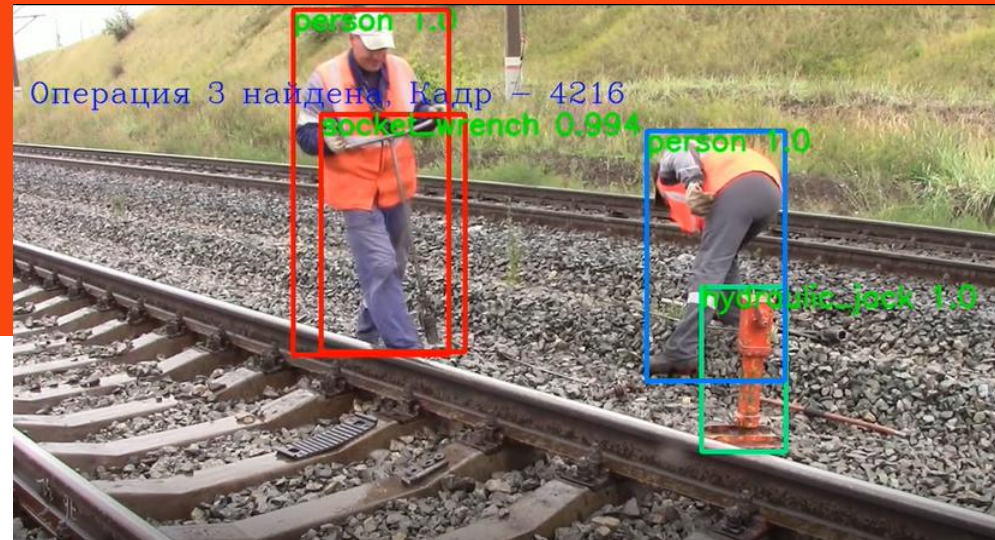
Keypoints



Operation detection

Operations with the set of

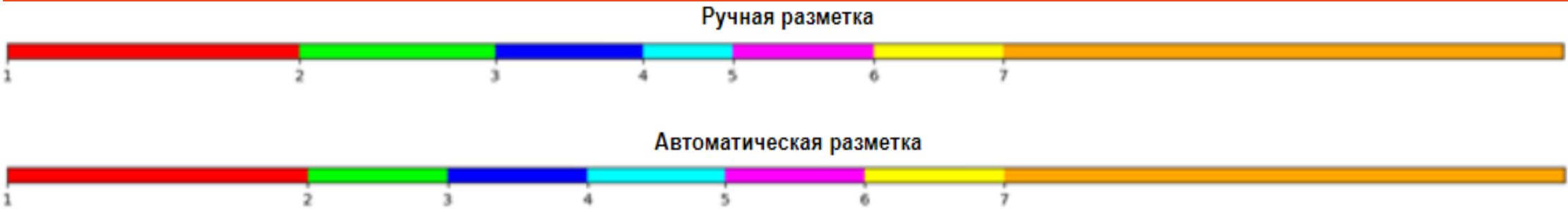
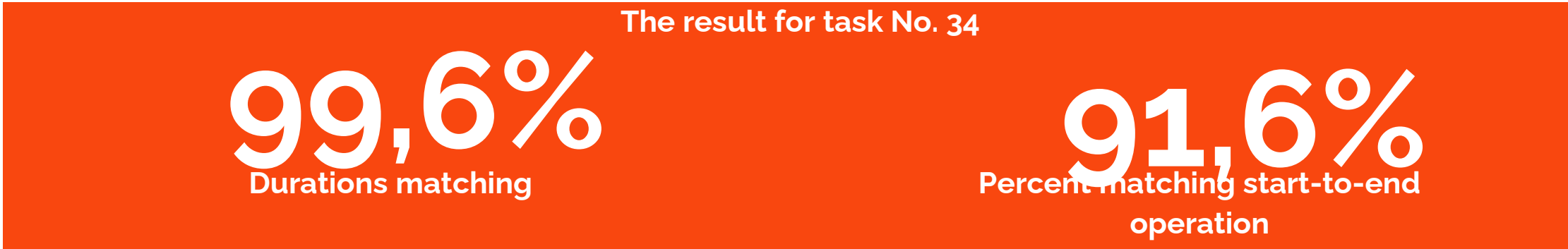
36
tools



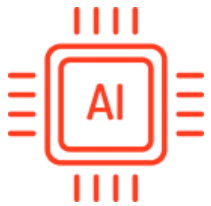
Business metrics of operations recognition

Comparison of markup by duration of operations, norm-min.

№ поручения	Ручная разметка	Автоматическая разметка	Комментарий
35	31,233	31,233	
34	39,433	39,600	
33	15,750	20,933	Несоответствие технологии операции (по норме - вилы, по факту - лопата)



Keypoints



Catboost
model



Binary
classification



Features - distances from the hands
keypoints to the instrument

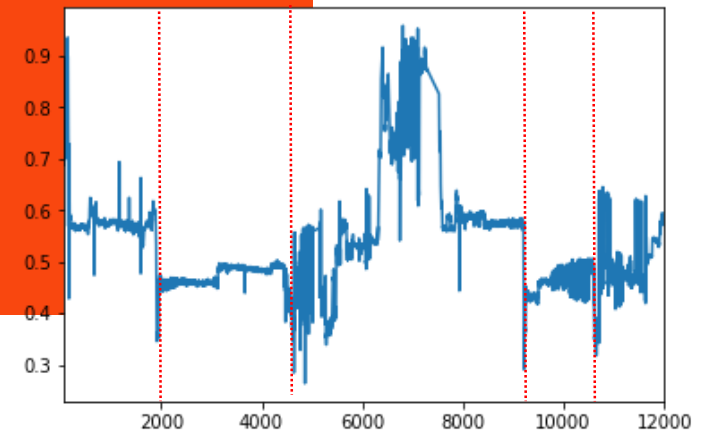
f1-score

0,83

Ключ (wrench):

Classification report for classifier model 1 video:

	precision	recall	f1-score	support
0.0	0.78	0.98	0.87	6878
1.0	0.96	0.63	0.76	5118
accuracy			0.83	11996
macro avg	0.87	0.81	0.82	11996
weighted avg	0.86	0.83	0.82	11996



Technological operations without tools



Human detection
(Employees)



Hands detection



Hands
keypoints



Operation
detection

122

Operations without
tools



Keypoints of the hands



3 coordinates



MediaPipe
format of
keypoints

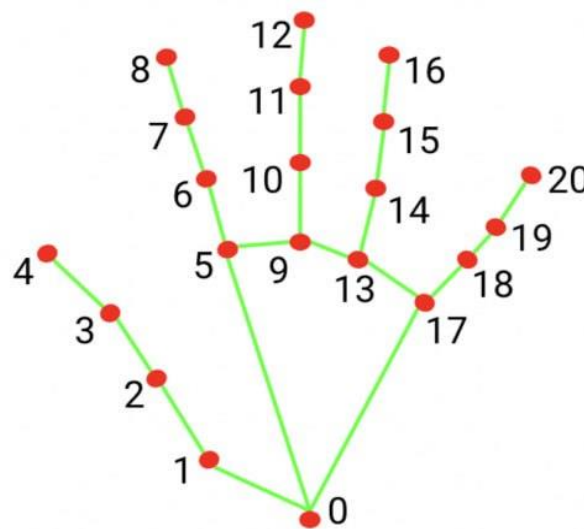


Normalized to 0..1

Description

20

keypoints



0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP

11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

Data: American Sign Language - phrases



Video phrases in
sign language



MediaPipe

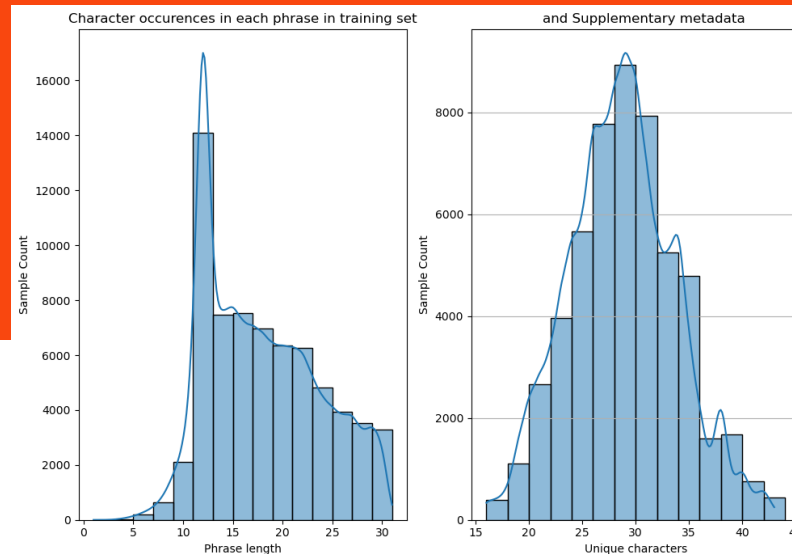


Sequence of keypoints
for each human
participant

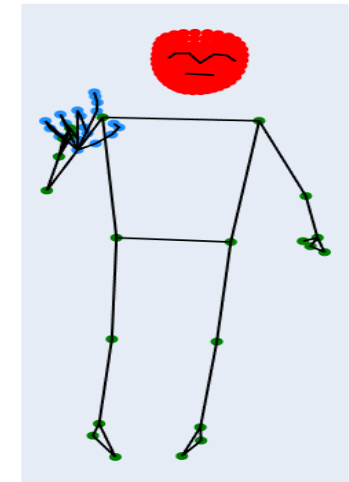
Data consists of

46478

phrases in ASL



Phrase: leona owens



Data. Sample test phrase

“lawanda”



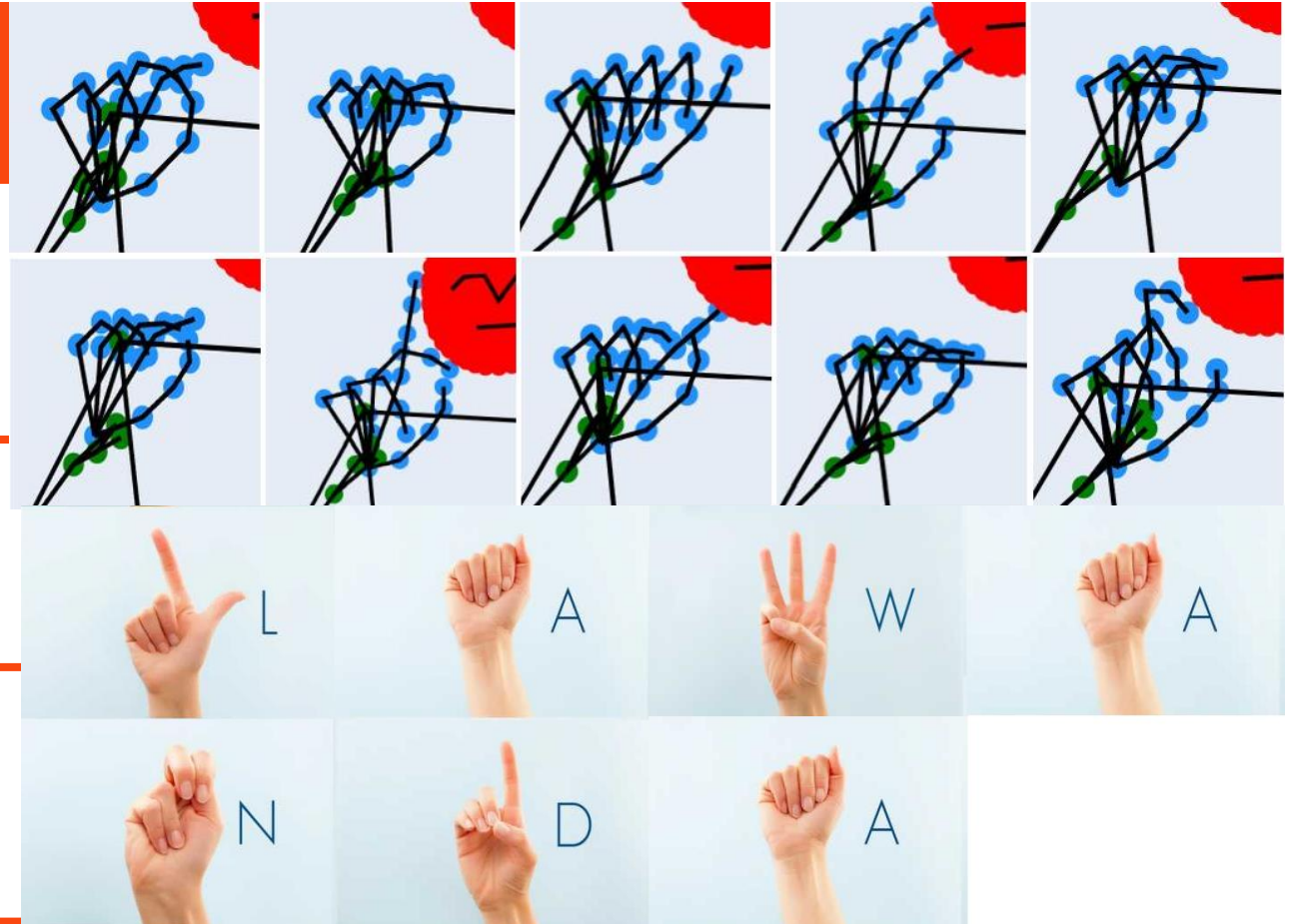
We have blanks in data



At the begining could
be any symbol



Hard to visually
recognize



Data. American Sign Language - Symbols



individual
characters dataset

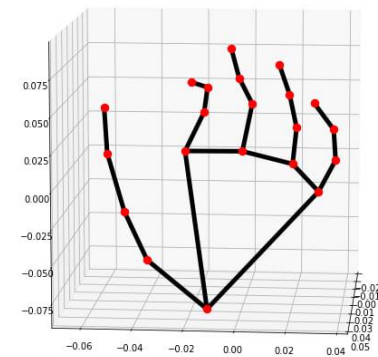


MediaPipe



Only hand keypoints -
 21×3

35
symbols in dataset



Data. Characters dataset distribution

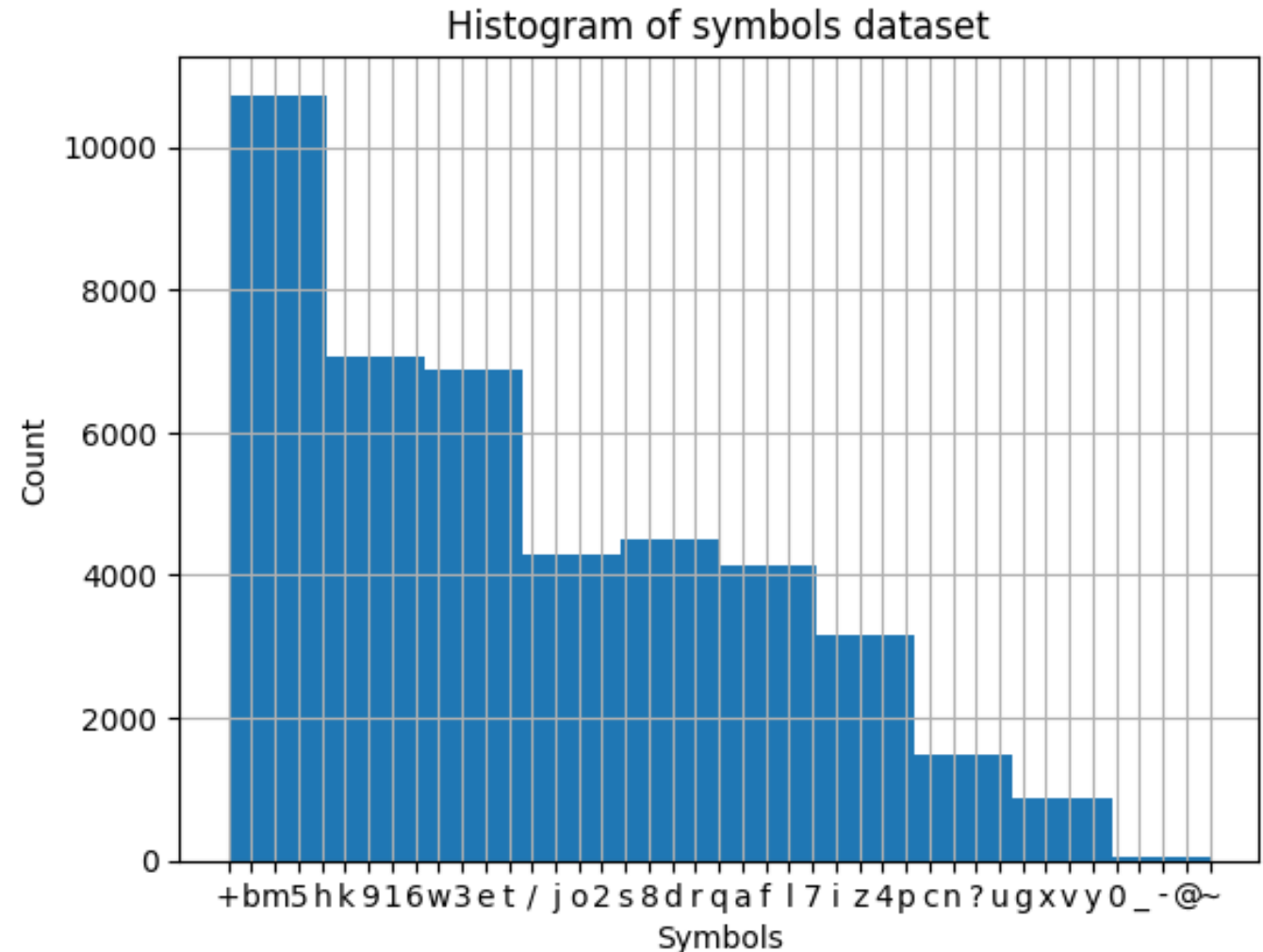


Dataset consists of 32 thousand

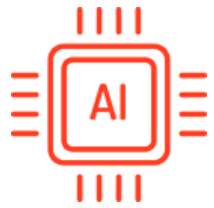
Training dataset 0.8 - 20 thousand

Validation dataset 0.2 - 6 thousand

Distribution of individual characters in dataset



Model. Feature selection



The distances between
the tips of the fingers



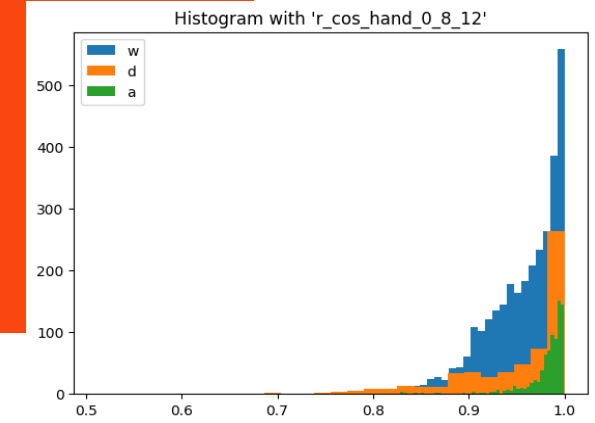
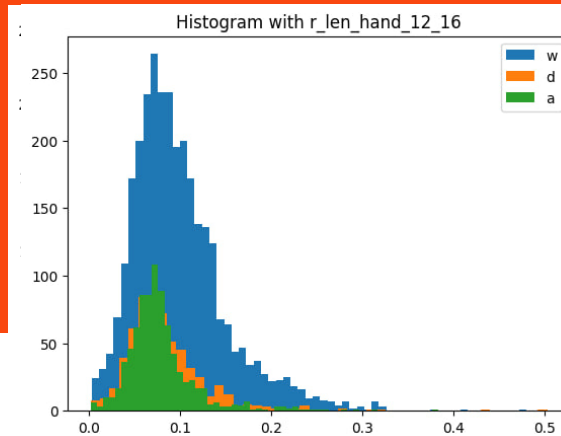
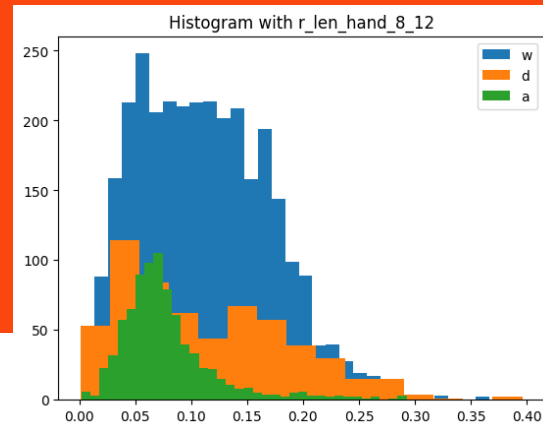
Fingers
length



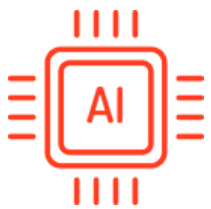
The cosine of the angle between
the wrist point and the rest of the
key points

Taking into
account

207
features



Model. Some results.



Catboost
method



Binary
classification

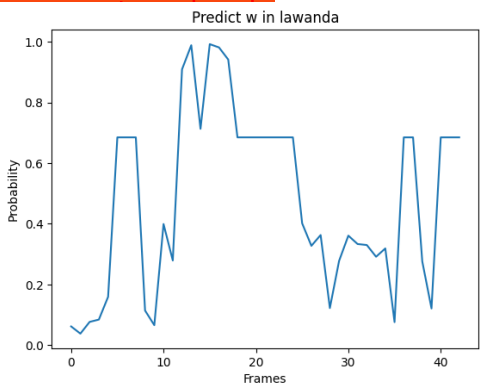


Classes w/abcd

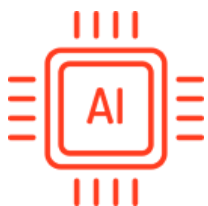
f1-score

0,79

	precision	recall	f1-score	support
0	0.74	0.74	0.74	185
1	0.83	0.83	0.83	276
accuracy			0.79	461
macro avg	0.78	0.78	0.78	461
weighted avg	0.79	0.79	0.79	461



Model. Results.



Catboost
method



Multiclass
classification

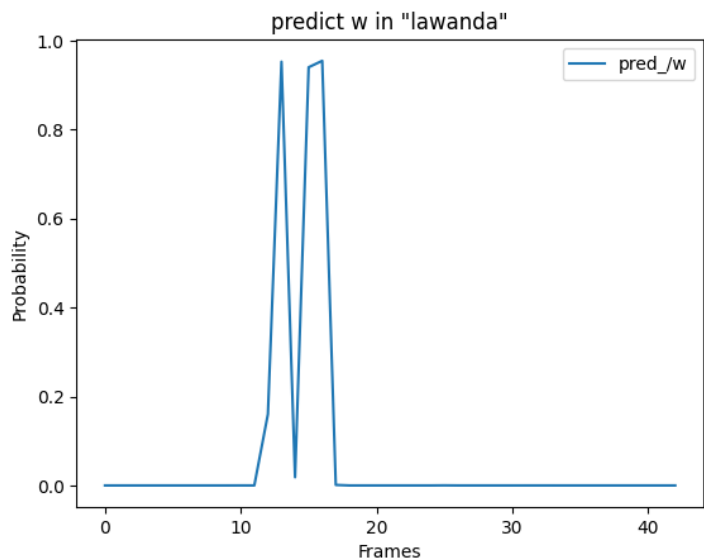


35 classes

f1-score

0,98

	precision	recall	f1-score	support
/	0.98	0.95	0.97	191
0	1.00	0.50	0.67	6
1	0.95	1.00	0.97	389
2	0.97	1.00	0.98	387
3	0.99	0.99	0.99	356
4	0.97	0.99	0.98	374
5	0.99	0.98	0.98	376
6	0.97	0.99	0.98	360
7	0.99	0.99	0.99	368
8	0.99	0.99	0.99	358
9	0.97	0.98	0.97	368
a	0.98	0.99	0.99	166
b	0.98	0.98	0.98	125
c	0.94	0.97	0.95	167
d	0.95	0.96	0.95	117
e	0.97	1.00	0.98	83
f	1.00	0.85	0.92	65
g	0.99	0.93	0.95	80
h	0.96	0.99	0.97	276
i	0.95	0.78	0.85	49
j	0.95	0.98	0.97	128
k	1.00	0.97	0.98	100
l	0.99	0.99	0.99	97
m	0.97	0.99	0.98	187
n	0.97	0.94	0.95	65
o	0.97	0.84	0.90	37
p	1.00	0.91	0.95	98
q	1.00	0.77	0.87	13
r	0.97	0.94	0.96	107
s	0.97	1.00	0.98	184
t	0.98	0.96	0.97	126
u	1.00	0.74	0.85	19
v	1.00	0.88	0.94	42
w	0.98	1.00	0.99	615
x	1.00	0.60	0.75	10
y	1.00	0.78	0.88	18
z	1.00	0.69	0.81	16
accuracy			0.98	6523
macro avg	0.98	0.91	0.94	6523
weighted avg	0.98	0.96	0.97	6523



Results. Recognition phase.

“lawanda”



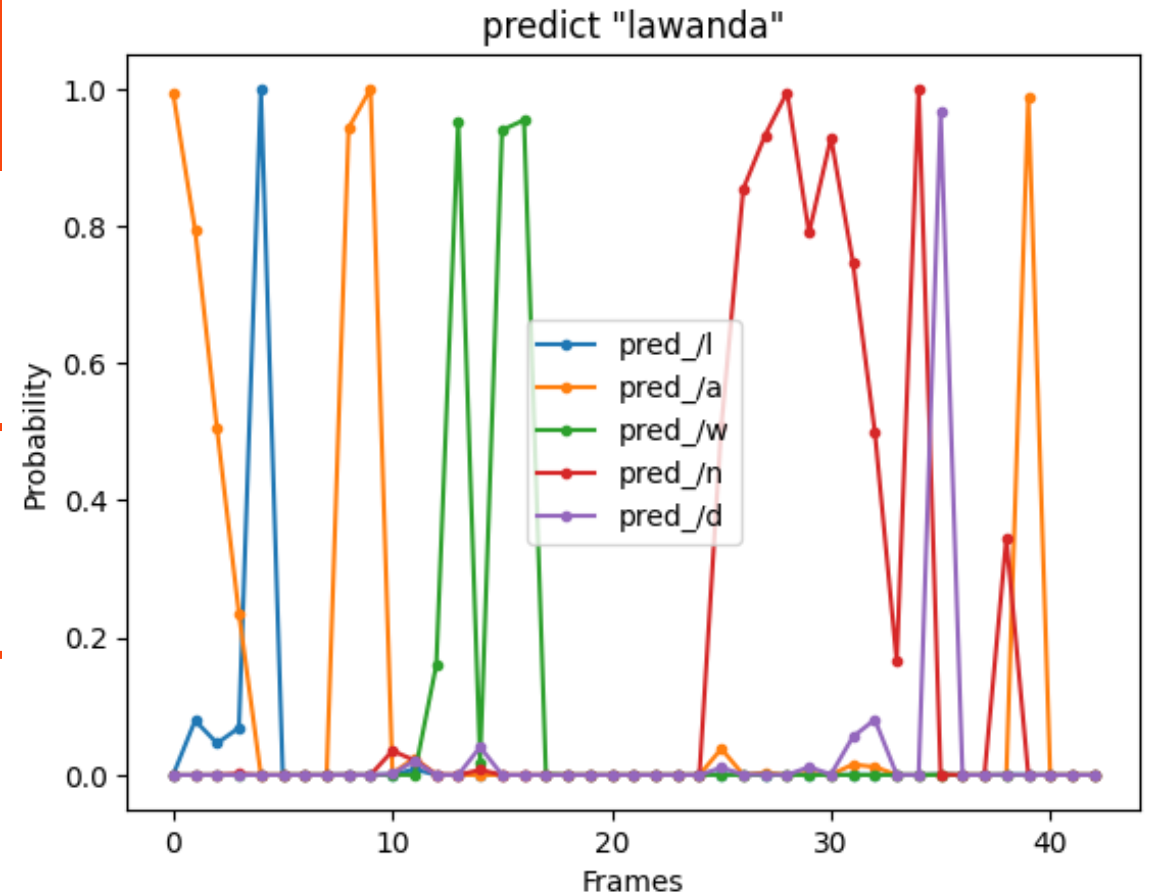
Multiclass classification



Threshold 0,95



CER = 0,29



Outcome. Recognition of key points

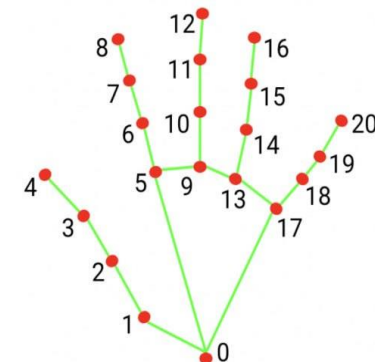
Keypoints



The approach has shown good results in gesture recognition



the approach is suitable for recognizing the sequence of finger movements in technological operations without tools



- | | |
|-----------------------|-----------------------|
| 0. WRIST | 11. MIDDLE_FINGER_DIP |
| 1. THUMB_CMC | 12. MIDDLE_FINGER_TIP |
| 2. THUMB_MCP | 13. RING_FINGER_MCP |
| 3. THUMB_IP | 14. RING_FINGER_PIP |
| 4. THUMB_TIP | 15. RING_FINGER_DIP |
| 5. INDEX_FINGER_MCP | 16. RING_FINGER_TIP |
| 6. INDEX_FINGER_PIP | 17. PINKY_MCP |
| 7. INDEX_FINGER_DIP | 18. PINKY_PIP |
| 8. INDEX_FINGER_TIP | 19. PINKY_DIP |
| 9. MIDDLE_FINGER_MCP | 20. PINKY_TIP |
| 10. MIDDLE_FINGER_PIP | |

Спасибо за внимание

Штехин Сергей Евгеньевич

Sergei.Shtekhin@OCR.V.RU

ОЦРВ

