

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Алтайский государственный технический университет им. И.И. Ползунова»

Разработка приложения для онлайн перевода с русского жестового языка

Автор:
студент группы ПИ-12, Функ Д.Е.

Научный руководитель:
доцент кафедры ПМ, к.ф.-м.н.
Андреева А.Ю.

Барнаул, 2025

Актуальность

- По данным ВОЗ около 470 миллионов человек страдают от потери слуха.
- К 2050 году количество таких людей достигнет 900 миллионов.
- Согласно статистике, каждый девятый житель Земли имеет нарушение слуха (около 650 миллионов), 70 из которых – глухие люди, более 200 миллионов живут с частичной потерей слуха.

Цель и задачи работы

Цель работы заключается в разработке приложения для онлайн перевода с русского жестового языка.

Поставленные задачи включают в себя:

- Подобрать датасет и провести предобработку данных;
- Создать алгоритм обучения нейросети для классификации жестов;
- Разработать алгоритм взаимодействия серверной части приложения с нейронной сетью;
- Разработать интерфейс для взаимодействия с разработанным программным обеспечением.

Набор данных

Датасет — SLOVO.

- 20 тысяч видео.
- 20 видео на один жест.
- 1000 классов жестов.
- Дополнительный класс «не жест».
- Датасет разбит на тренировочную (train), и тестовую (test) выборки с 15 000 и 5 000 сэмплами соответственно.

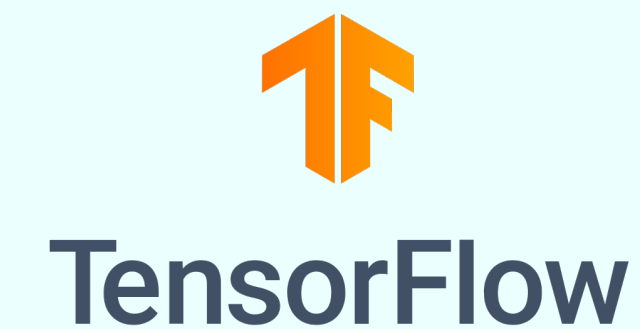


Пример жеста «в восемь пятнадцать»

Средства разработки

IDE Обработка данных Выделение Обучение Приложение

КЛЮЧЕВЫХ ТОЧЕК



PyCharm
2024.2.3
(Professional
Edition)

Архитектура модели

Входные данные

- Что: Координаты ключевых точек тела (губы, руки, поза) из каждого кадра видео.
- Формат: Тензор $[N_{\text{кадров}} \times N_{\text{точек}} \times (x, y, z)]$.

Предобработка

- Устранение пропущенных значений;
- Нормализация данных(Z-score).

$$x_i = \begin{cases} 0, & \text{если } x_i = NaN \\ x_i, & \text{иначе.} \end{cases}$$

где x_i значение из набора данных.

$$x_{\text{standardized}} = \frac{x - \mu}{\sigma}$$

где $x_{\text{standardized}}$ нормализованное значение;

x значение координаты;

μ — среднее значение признака;

σ — стандартное отклонение признака.

$$\mu_j = \frac{1}{N} \sum_{i=1}^N x_{ij}$$

где μ_j — среднее значение j -го признака;

N — количество наблюдений;

x_{ij} — значение j -й точки в i -м кадре;

значение из набора данных.

$$\sigma_j = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_{ij} - \mu_j)^2}$$

где σ_j стандартное

отклонение j -го признака;

Архитектура модели

Генерация эмбеддингов

- CustomEmbedding:
 - Преобразует координаты в вектор фиксированной размерности.
 - Учитывает:
 - Пропущенные точки.
 - Важность групп точек (обучаемые веса).
 - Порядок кадров.

Временная обработка

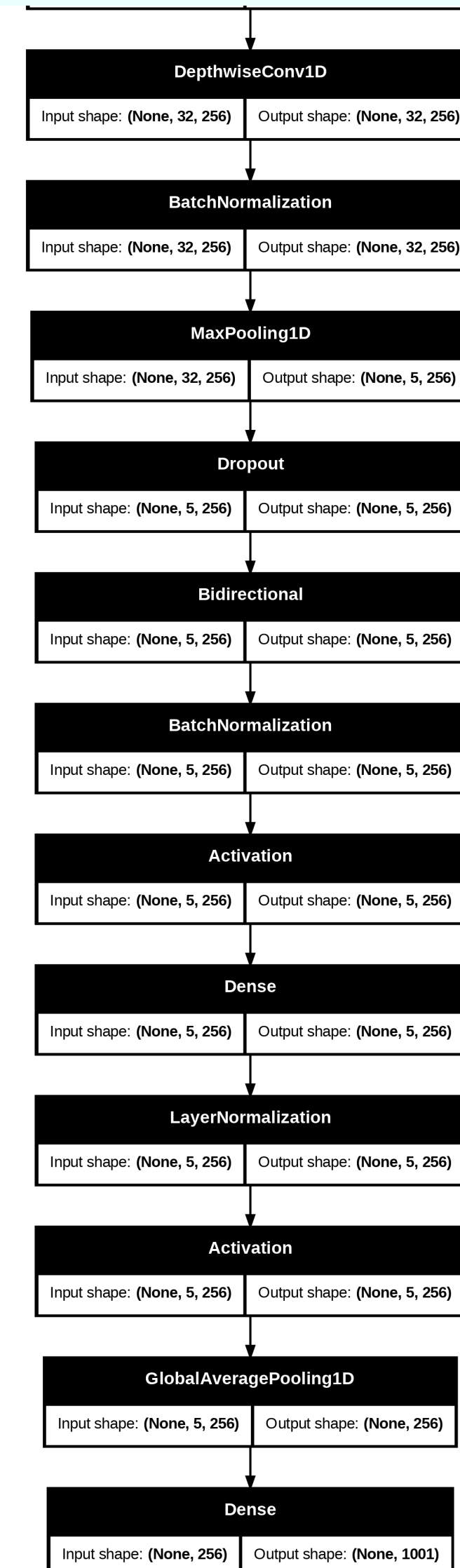
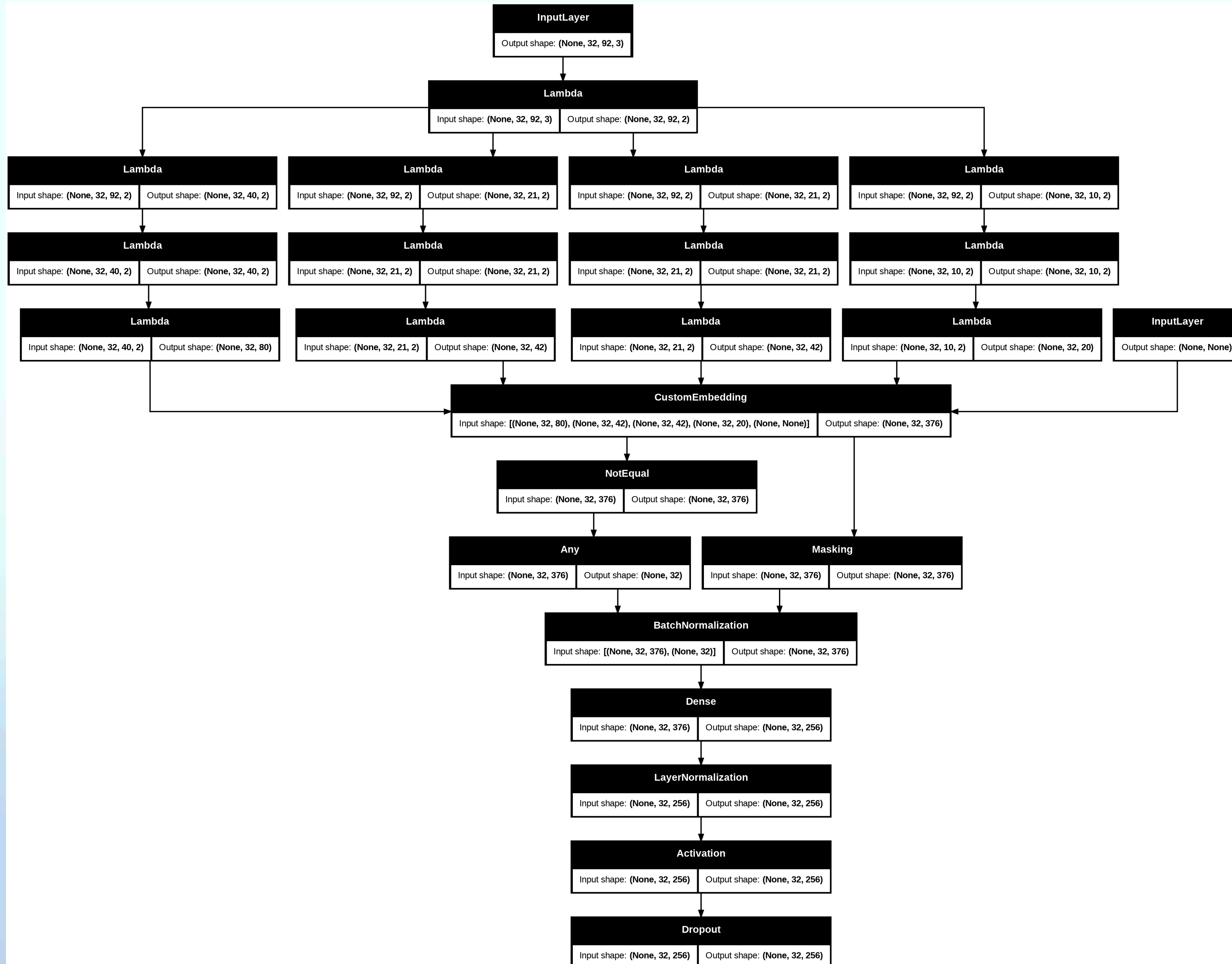
- Свертки 1D + LSTM → анализ последовательностей.
- MaxPooling → уменьшение размерности.

Классификация:

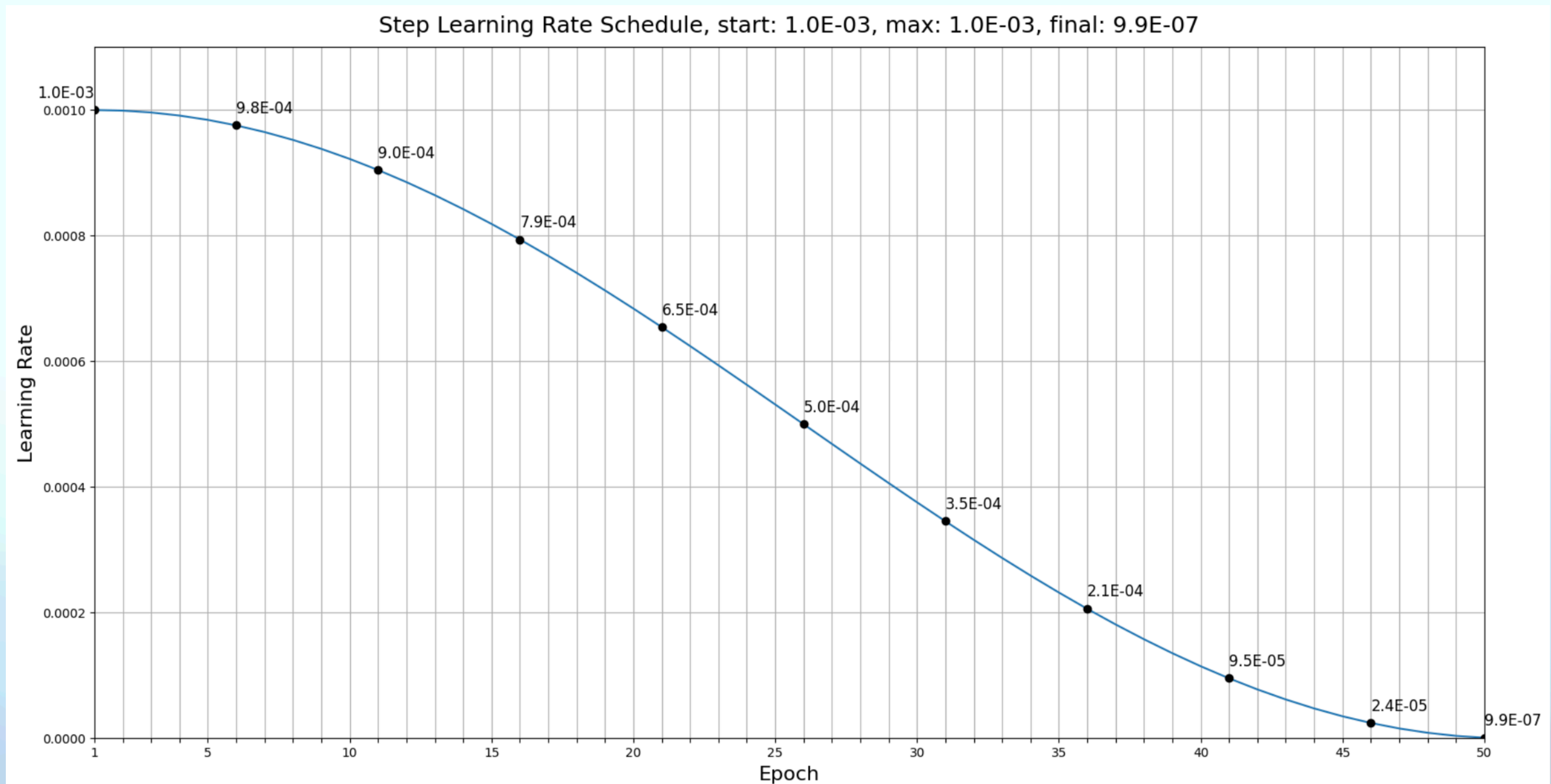
- GlobalAveragePooling → усреднение по времени.
- Dense-слой с softmax → прогноз класса жеста.

Архитектура модели

$$Loss = -\frac{1}{N} \sum_{i=1}^N \log(p_i)$$

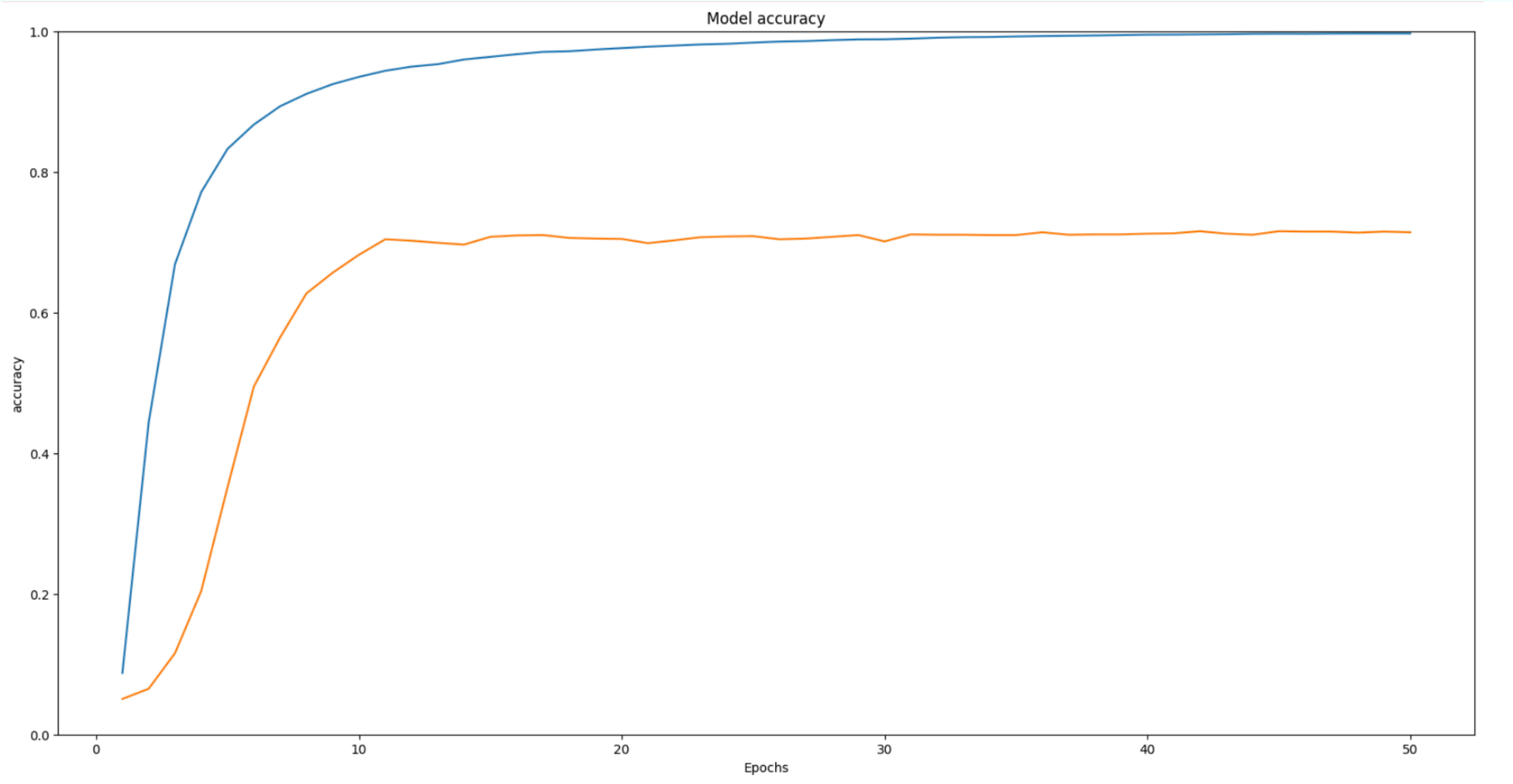


Обучение



Точность

$$accuracy = \frac{TP + TN}{FP + FN + TP + TN},$$



	Unnamed: 0	precision	recall	f1-score	text	pred_type
0	2_часа	1.00	1.00	1.00	2_часа	phrase
1	красивый	1.00	1.00	1.00	красивый	word
2	навсегда	1.00	1.00	1.00	навсегда	word
3	наблюдательность	1.00	1.00	1.00	наблюдательность	word
4	мятежный	1.00	1.00	1.00	мятежный	word
...	...	...	...	...	...	...

Основные результаты

```
metrics.accuracy_score(y_test, y_test_pred)
```

0.9634

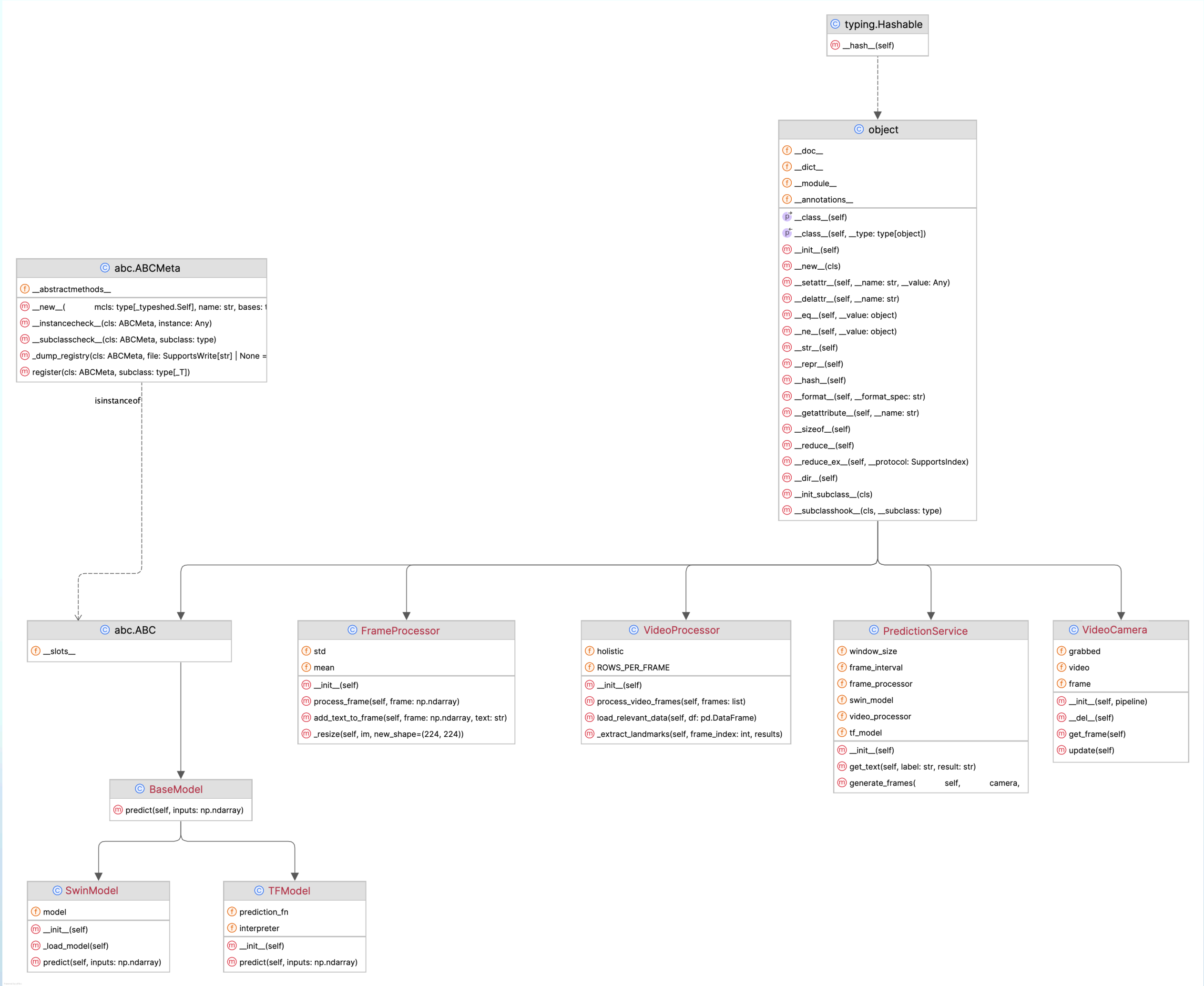
Достигнута точность классификации на уровне 96% (на тестовом наборе данных);

$$precision = \frac{TP}{TP + FP} \quad recall = \frac{TP}{TP + FN}$$

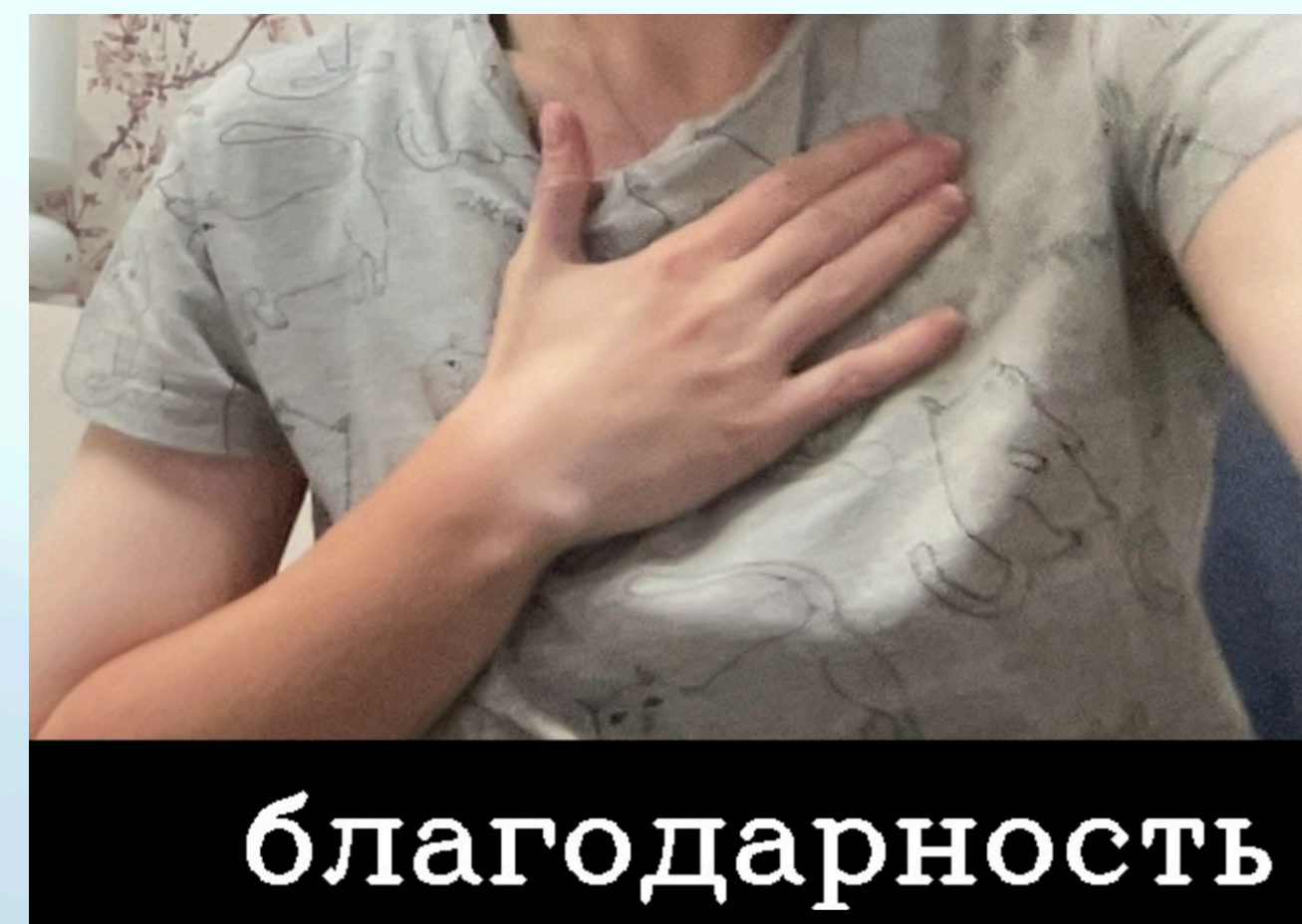
$$f_1 = 2 \times \frac{precision \times recall}{precision + recall}$$

где TP — true positive: классификатор верно отнёс объект к рассматриваемому классу;
TN — true negative: классификатор верно утверждает, что объект не принадлежит к рассматриваемому классу;
FP — false positive: классификатор неверно отнёс объект к рассматриваемому классу;
FN — false negative: классификатор неверно утверждает, что объект не принадлежит к рассматриваемому классу.

UML-диаграмма классов программы



Интерфейс



Заключение

- Выбран качественный датасет;
- Разработана архитектура нейронной сети, сочетающая пространственную (эмбеддинги ключевых точек) и временную (LSTM, свертки 1D) обработку данных;
- Реализованы механизмы маскирования кадров и регуляризации, улучшившие обобщающую способность модели;
- Достигнута точность классификации на уровне 96% (на тестовом наборе данных);
- Спроектирован интерфейс приложения;
- Исходный код размещён в репозитории Github.