

# Автоматическое распознавание речи

Модель keyword spotting.

Сигорова Анна Павловна

# Table of Contents.-

**01** - О задаче

**02** - О датасете

**03** - Выбор слова-триггера

**04** - Анализ датасета

**05** - Аугментация

**06** - Метрики

**07** - MatchboxNet

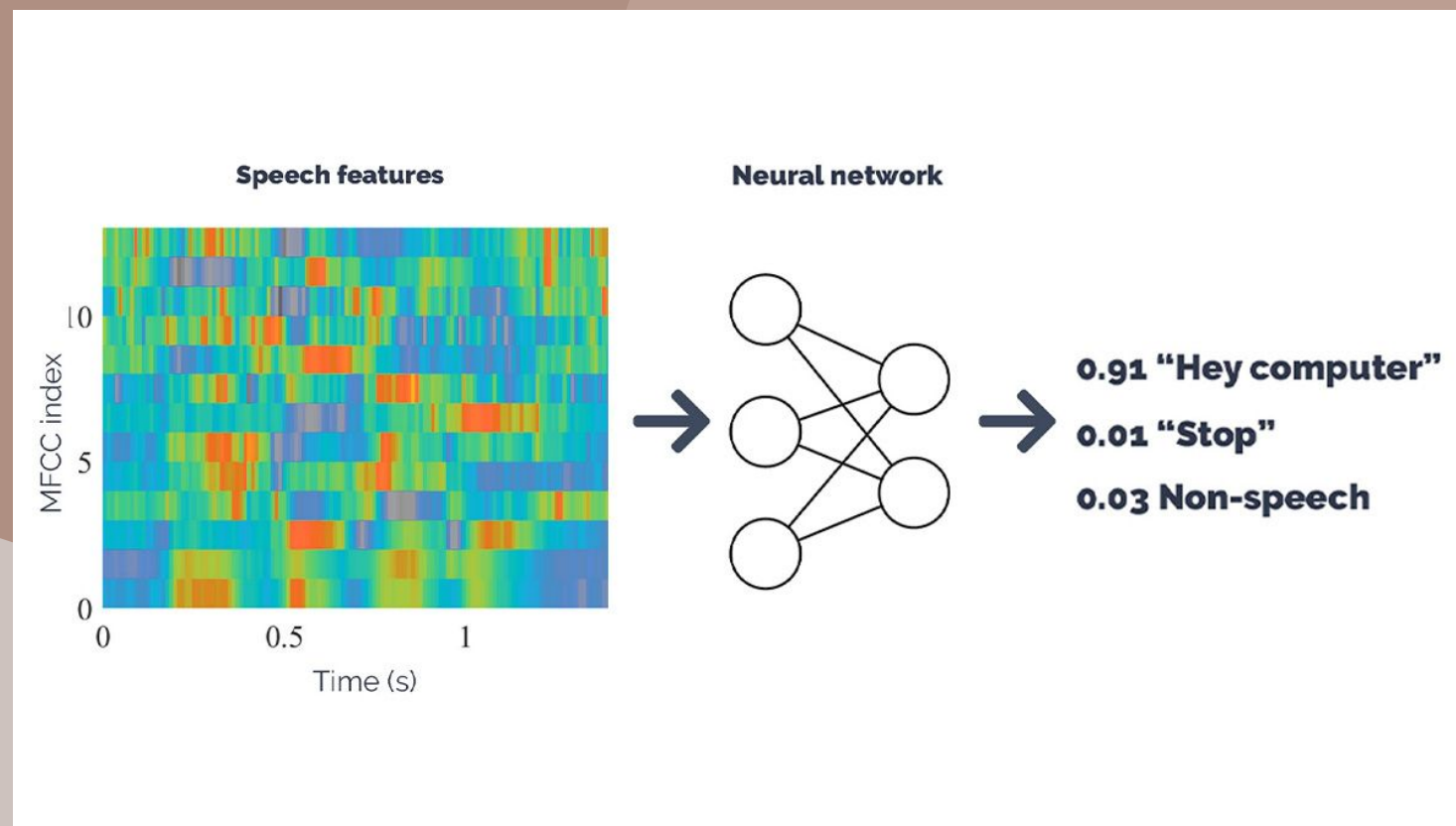
**08** - Эксперименты

**09** - Задумка

# О задаче keyword spotting

Задача классификации конкретного слова или выражения.

# 01



# 02

## О датасете

Был использован - Google speech command datasets v2.

Версия 2 была выпущена 11 апреля 2018 года и содержит 105 829 аудиофайлов.

	train	validation	test
v0.02	84848	9982	4890

# 03

## Выбор слова-триггера

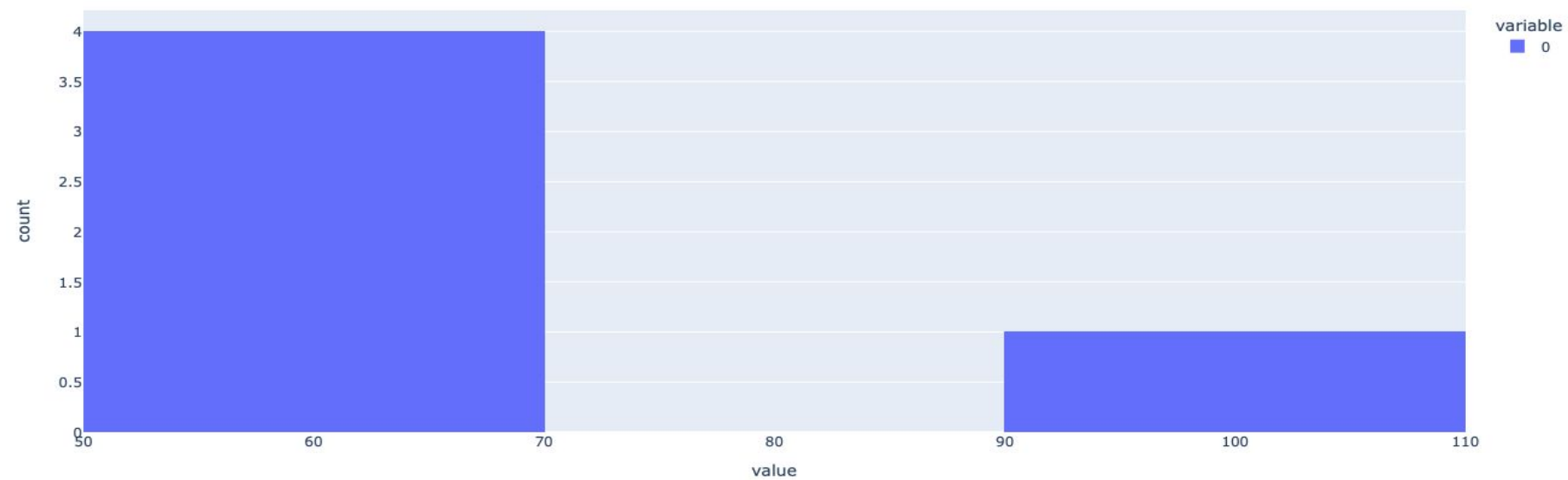
	train	validation	test
Общее	84848	9982	4890
Marvin	1710	195	11
Sheila	1606	204	17

The background features a mix of muted brown and beige tones with large, flowing abstract shapes. In the top-left corner, there is a cluster of small, dark brown dots. In the bottom-right corner, there is a cluster of small, light beige dots. A thin, wavy line in a slightly darker brown color runs across the top-right and bottom-left areas.

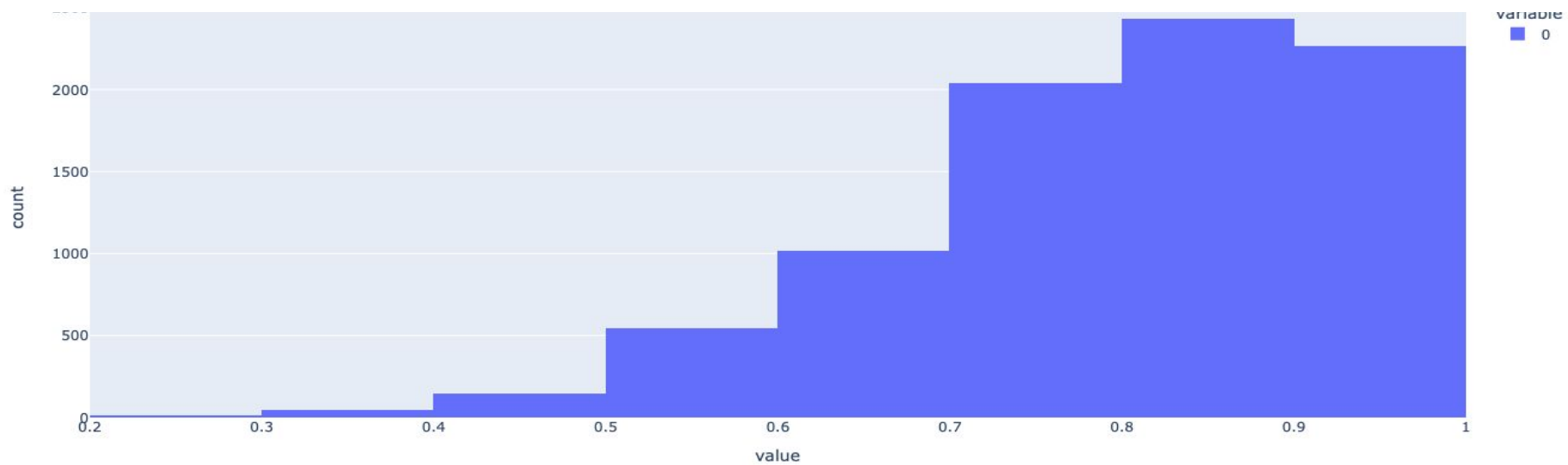
04

Анализ датасета

# Обучающая

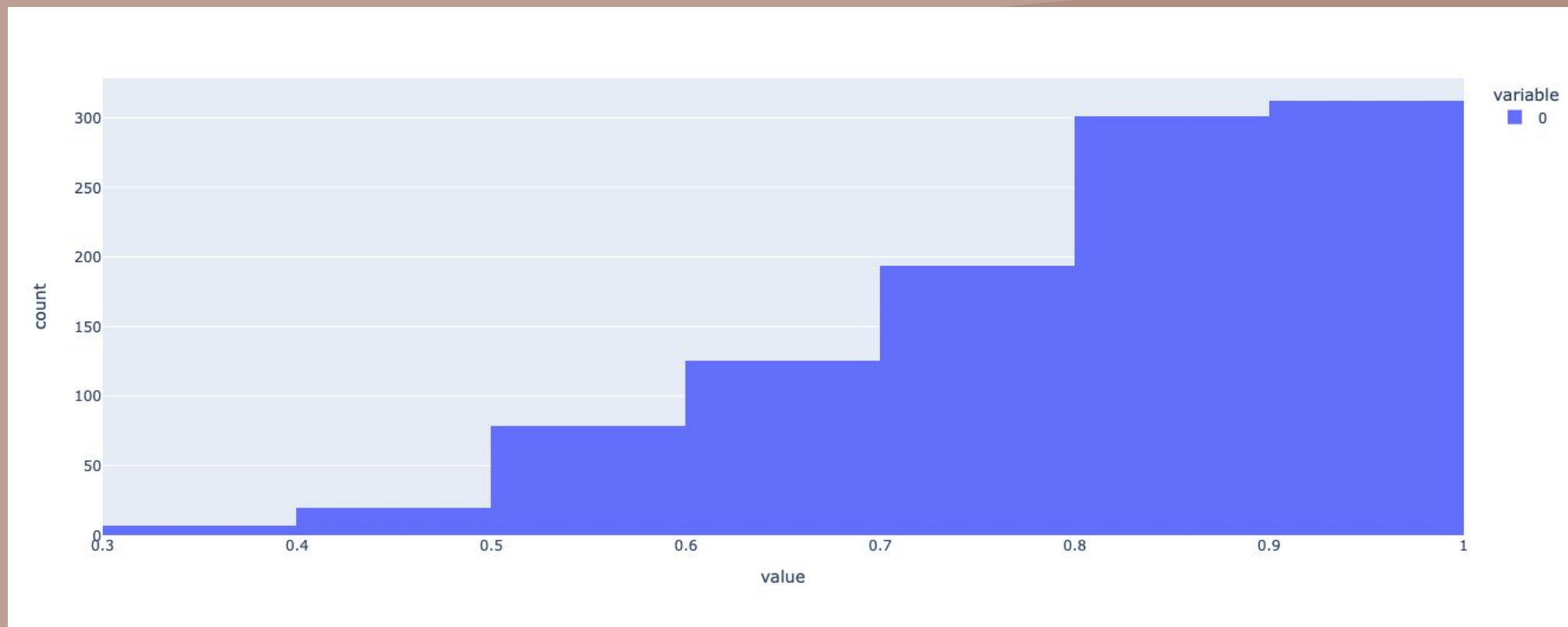


$t > 1$



$t < 1$

# Валидационная



$t < 1$



# 05

## Аугментация

Были применены следующие методы:

- Небольшое замедление голоса и его тона
- Реверберация
- Белый шум
- Добавление звуков на заднем фоне

# 06

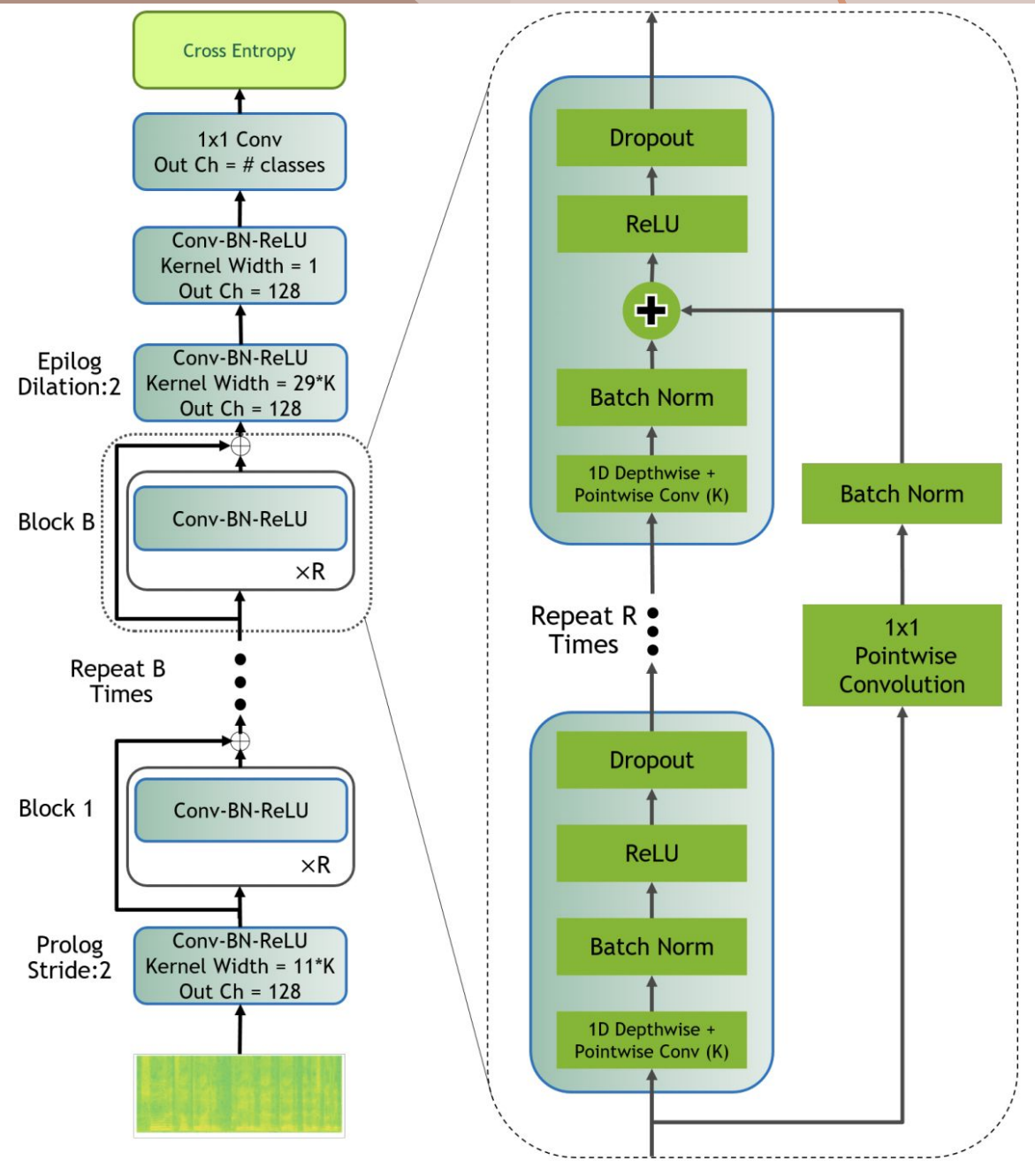
## Метрики

Были использованы следующие метрики:

- **Accuracy**
- **Balanced accuracy**
- **F1-score**
- **Far per hour**

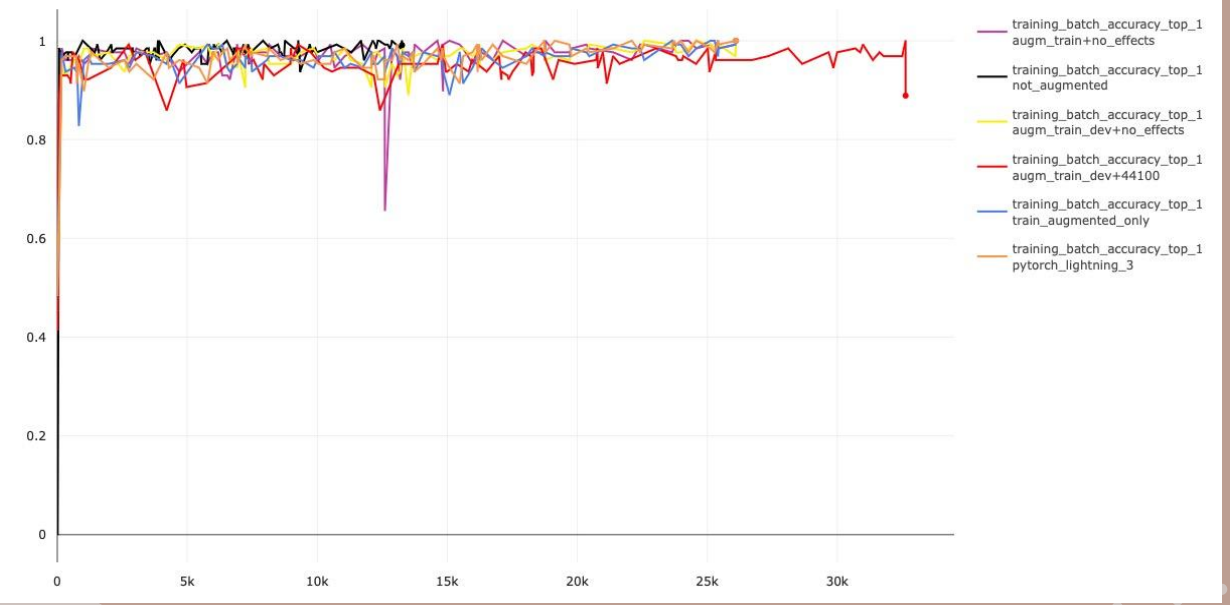
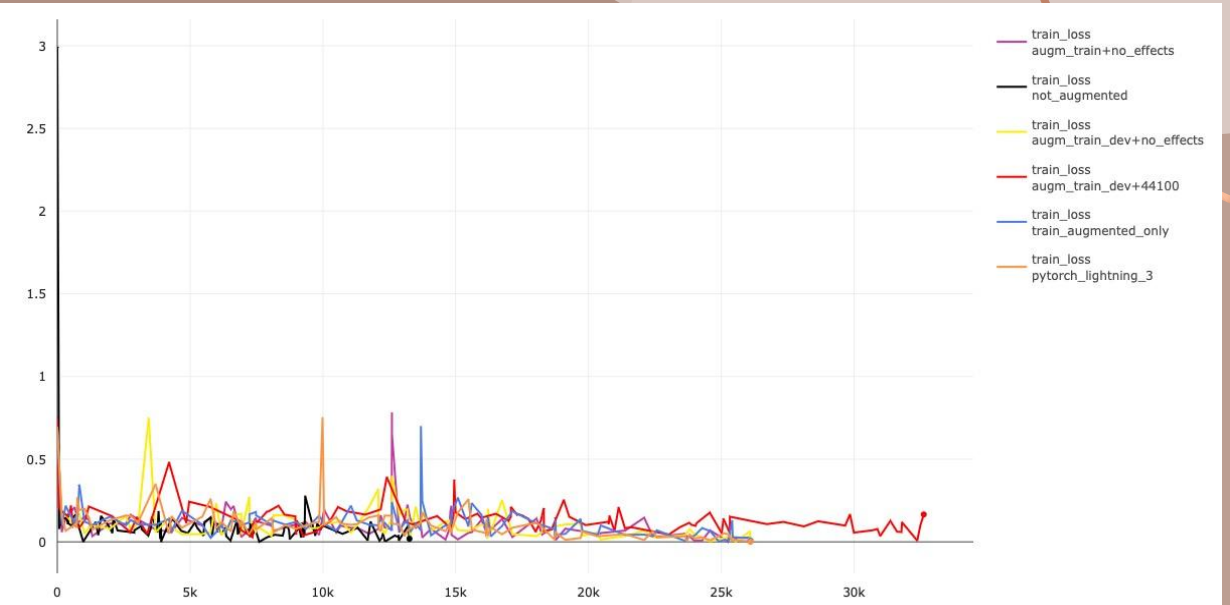
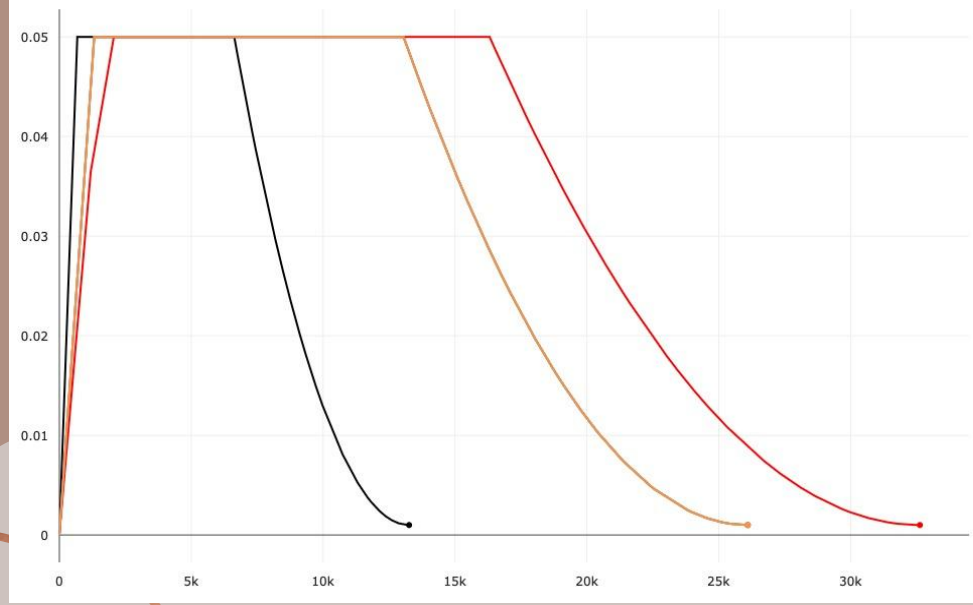
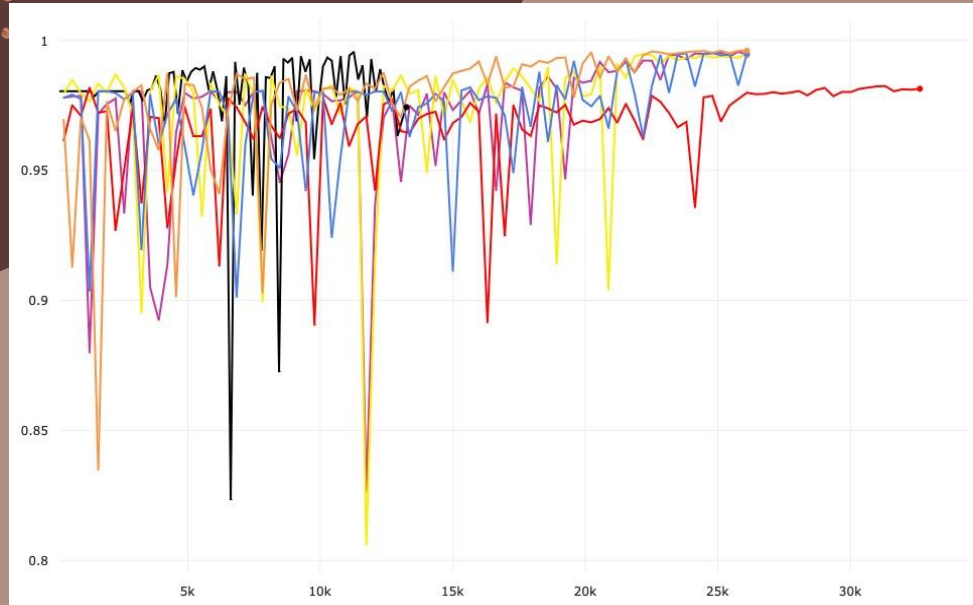
# 07

## MatchboxNet



08

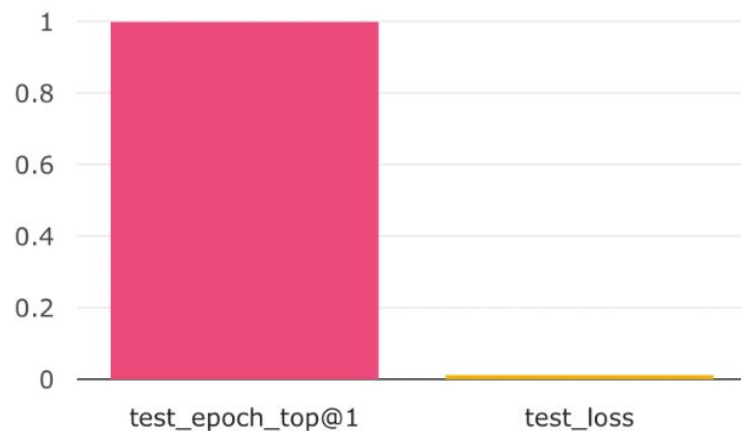
Эксперименты



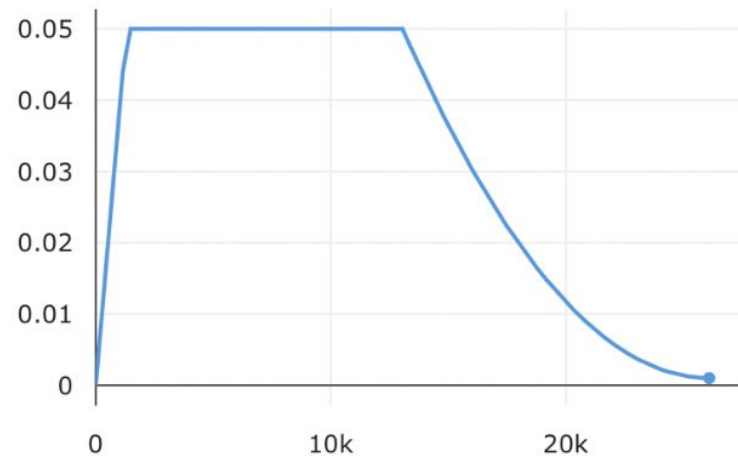
		Accuracy	Balanced Accuracy	F1-score	FAR per hour
Exp1 train + dev augm	train	99.85	99.85	99.85	81
	dev	99.61	99.61	99.62	
	test	99.84	90.85	99.92	
Exp2 only train augm	train	99.6	99.6	99.6	43
	dev	99.51	87.94	99.75	
	test	99.88	90.87	99.94	
Exp3 sr=44100	train	98.67	98.67	98.68	0
	dev	98.19	98.19	98.21	
	test	99.39	54.34	99.69	
Exp4 No effects All augm	train	99.65	99.65	99.65	3460
	dev	99.67	99.43	99.43	
	test	99.43	86.23	99.84	
Exp5 Not augmented	train	99.48	91.46	99.73	13
	dev	99.84	91.22	99.77	
	test	99.54	81.78	99.92	
Exp6 No effects train augm	train	99.7	99.7	99.7	3534
	dev	99.73	90.72	99.77	
	test	99.56	86.26	99.87	

# Эксперимент 1

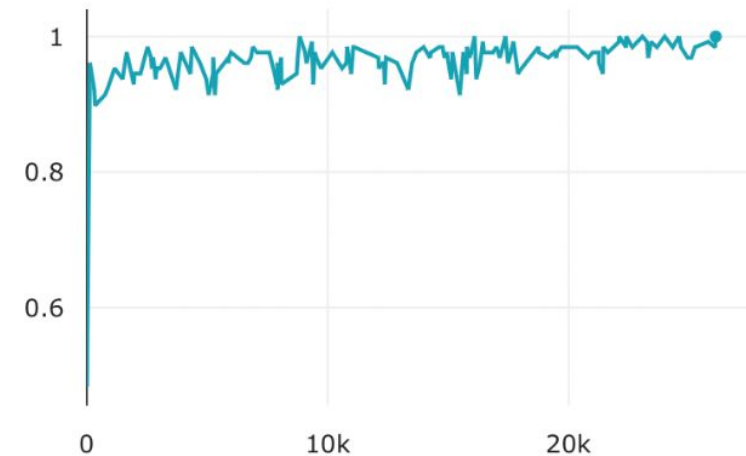
test\_epoch\_top@1, test\_loss



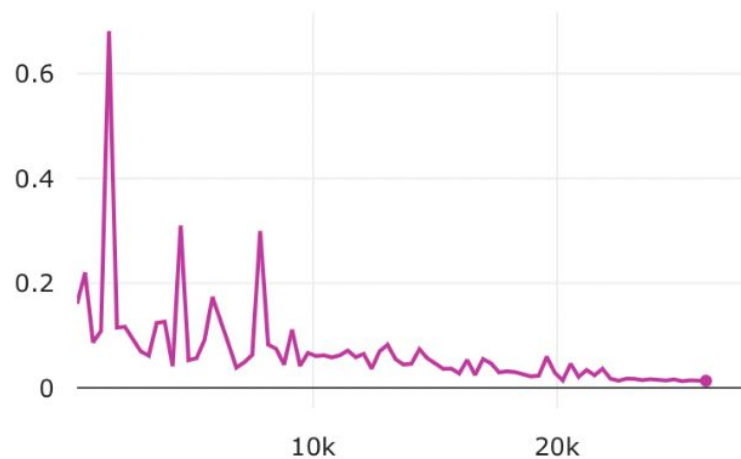
learning\_rate



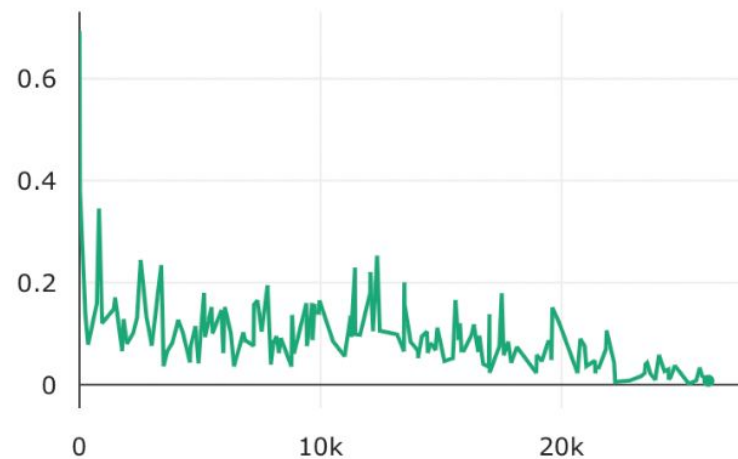
training\_batch\_accuracy\_top\_1



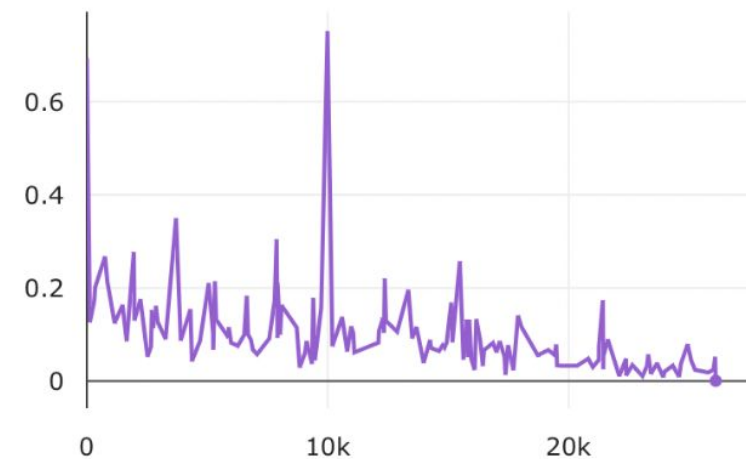
val\_loss



loss



train\_loss



# Эксперимент 1

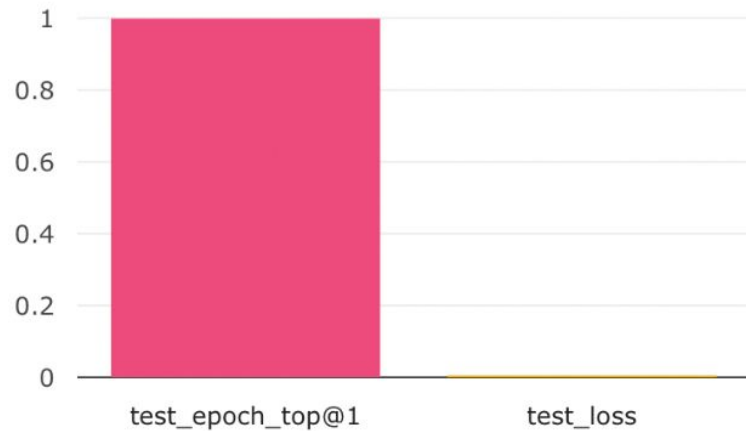
## Marvin



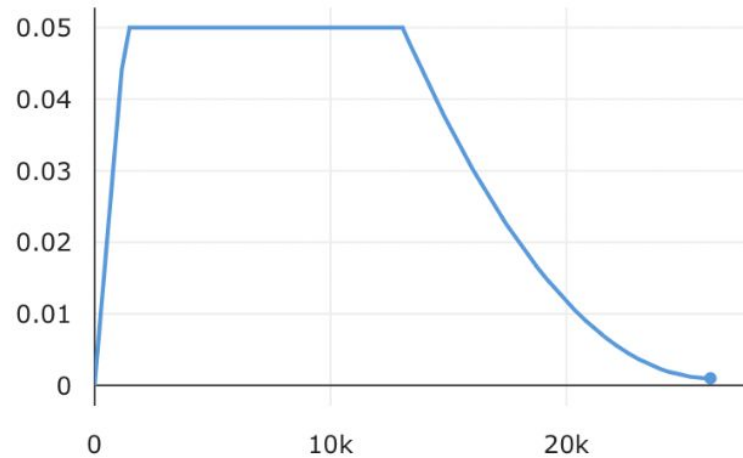


# Эксперимент 2

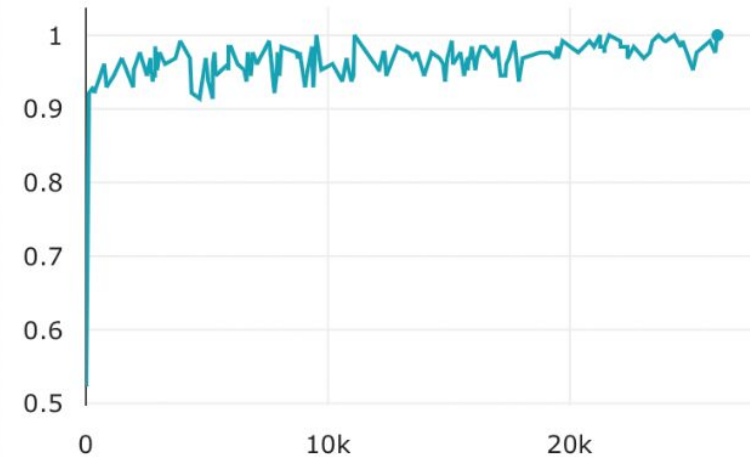
test\_epoch\_top@1, test\_loss



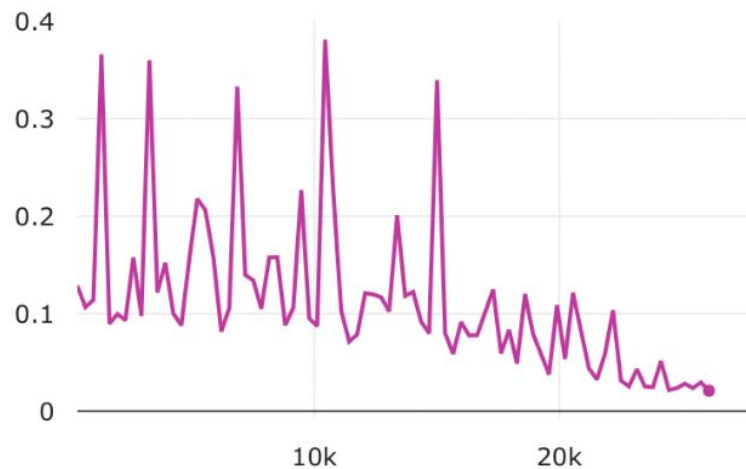
learning\_rate



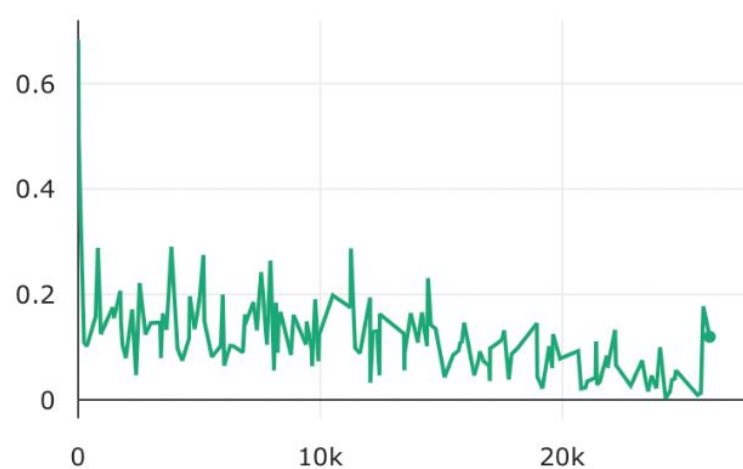
training\_batch\_accuracy\_top\_1



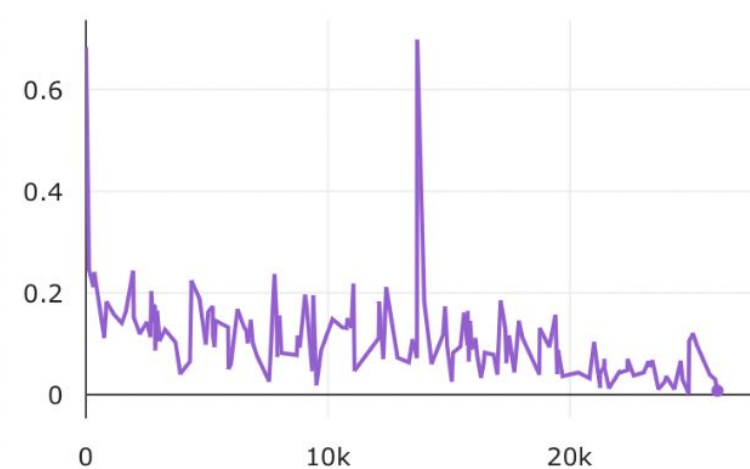
val\_loss



loss

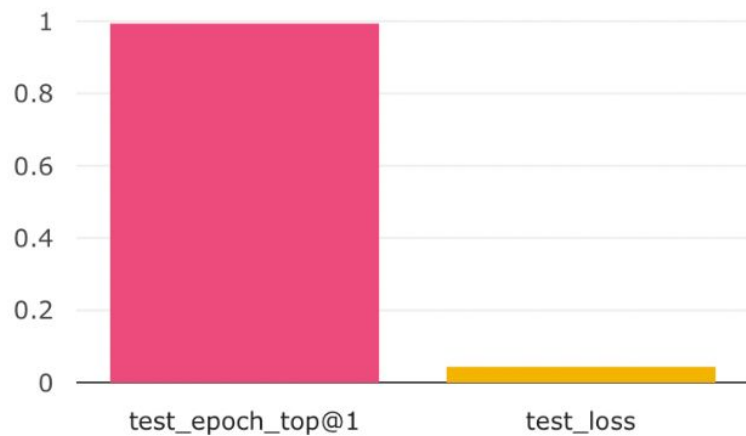


train\_loss

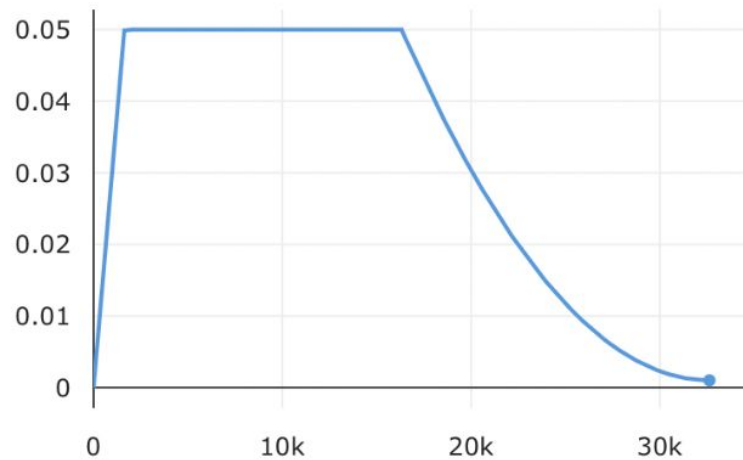


# Эксперимент 3

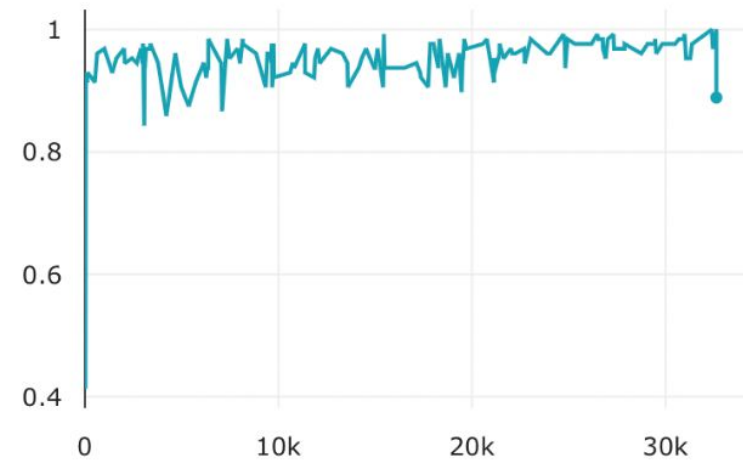
test\_epoch\_top@1, test\_loss



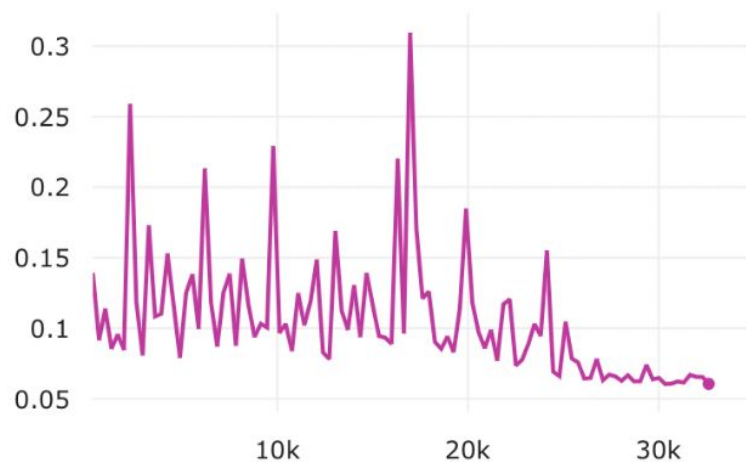
learning\_rate



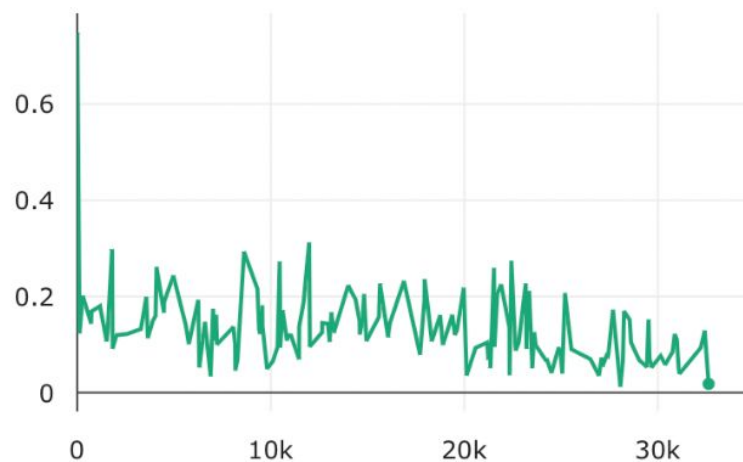
training\_batch\_accuracy\_top\_1



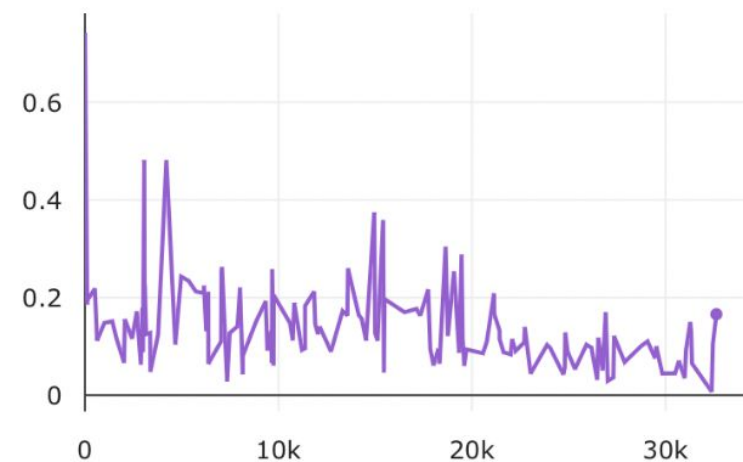
val\_loss



loss

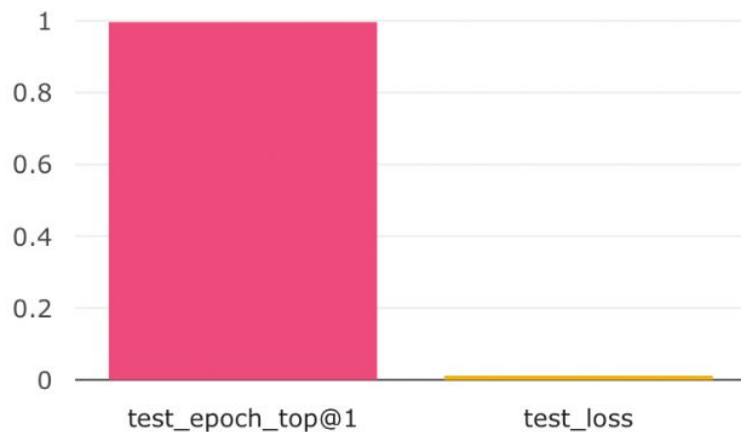


train\_loss

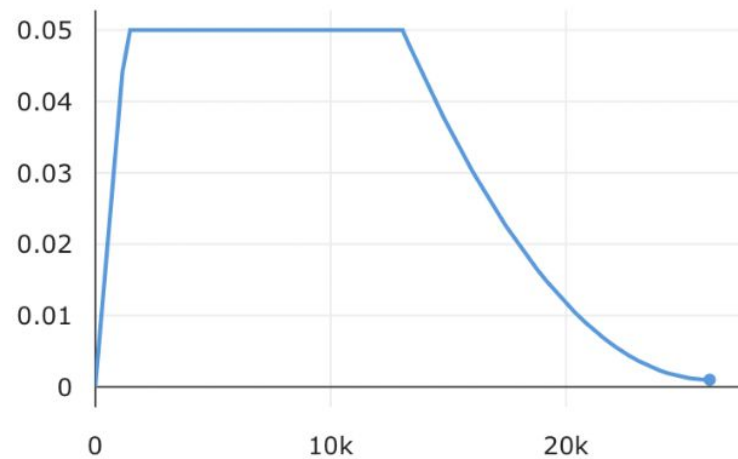


# Эксперимент 4

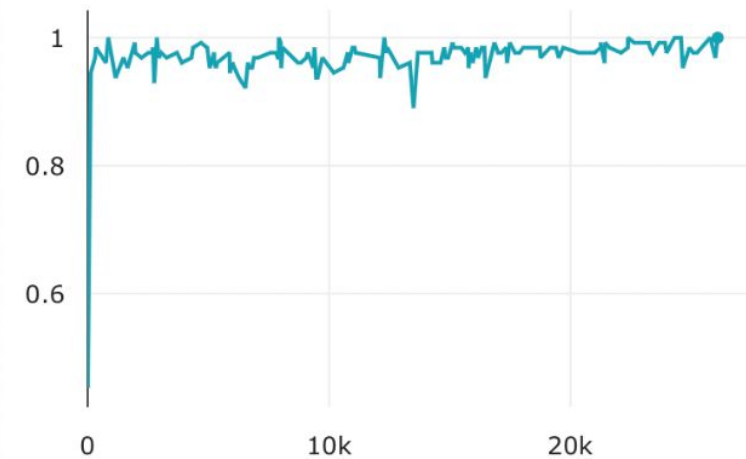
test\_epoch\_top@1, test\_loss



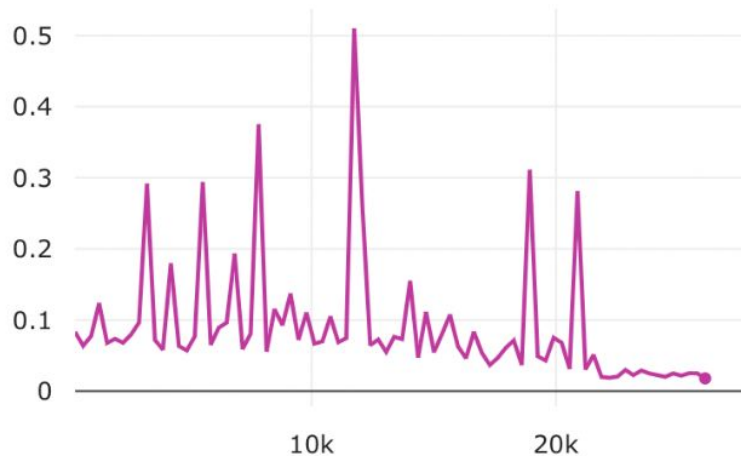
learning\_rate



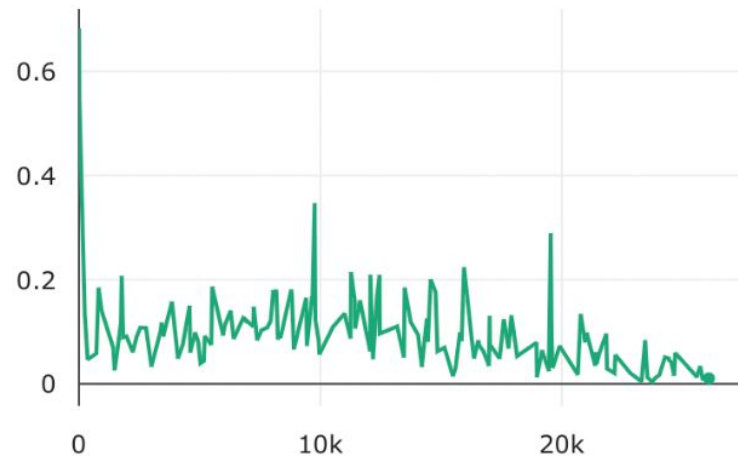
training\_batch\_accuracy\_top\_1



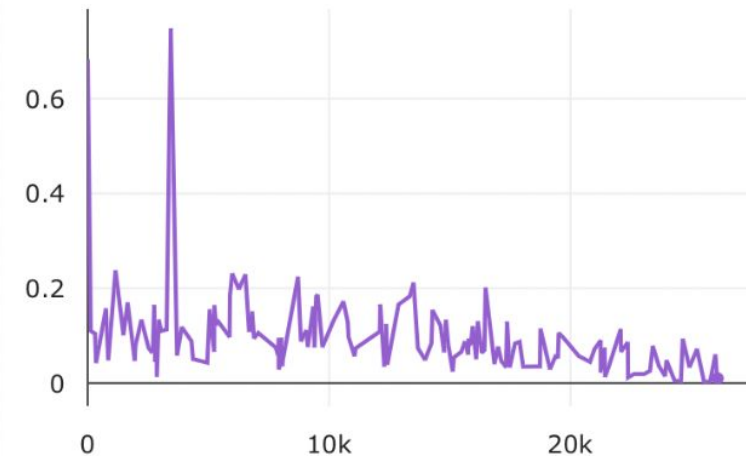
val\_loss



loss

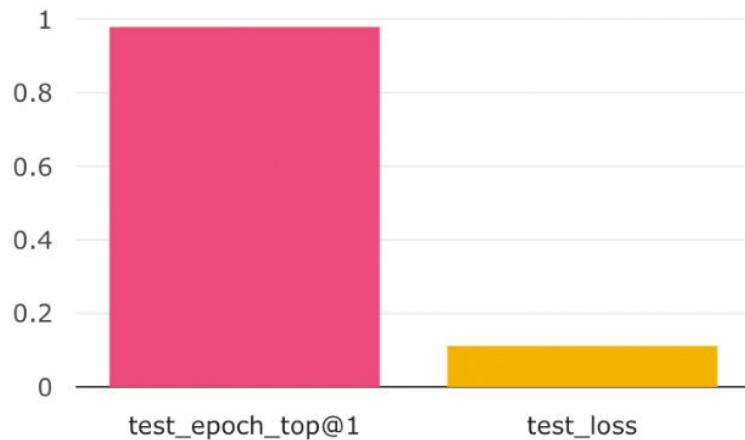


train\_loss

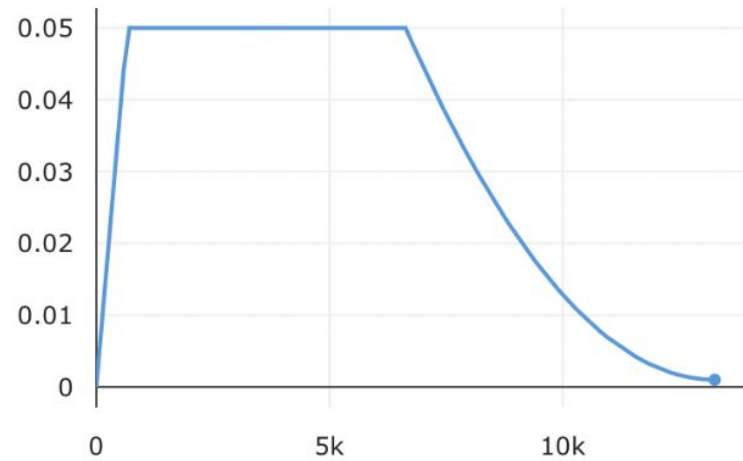


# Эксперимент 5

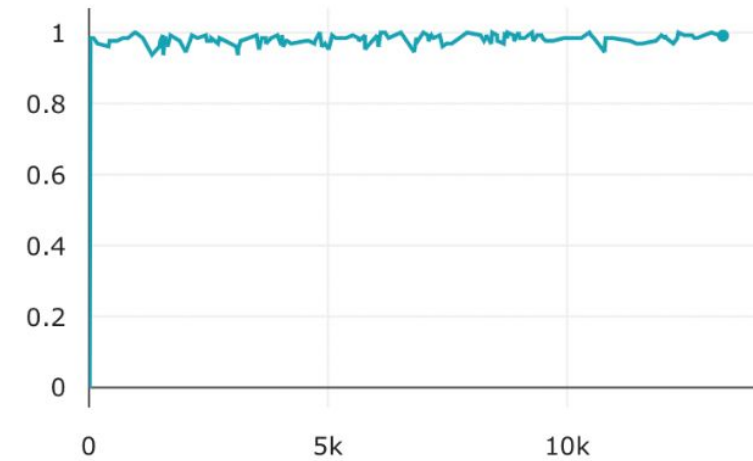
test\_epoch\_top@1, test\_loss



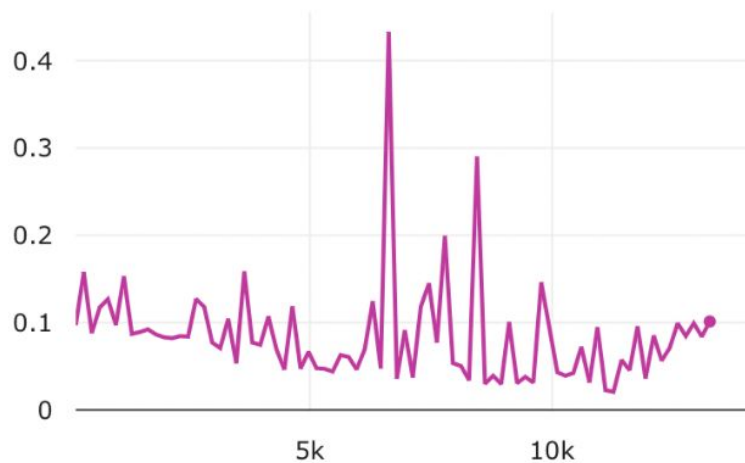
learning\_rate



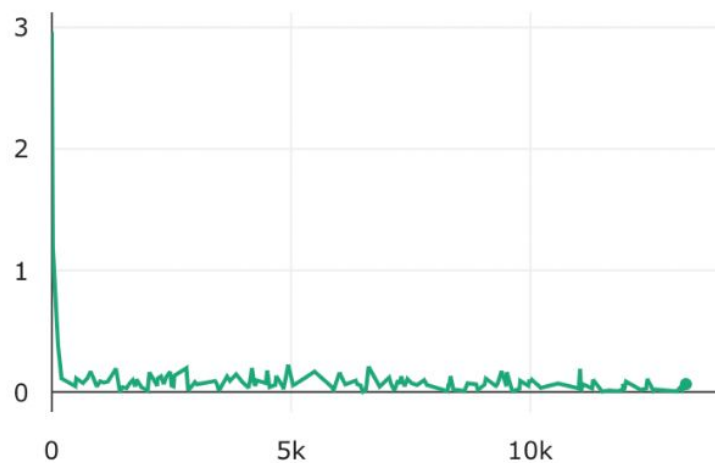
training\_batch\_accuracy\_top\_1



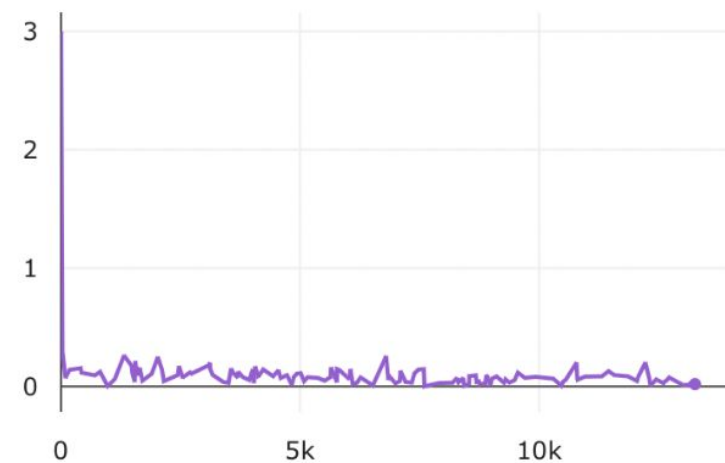
val\_loss



loss

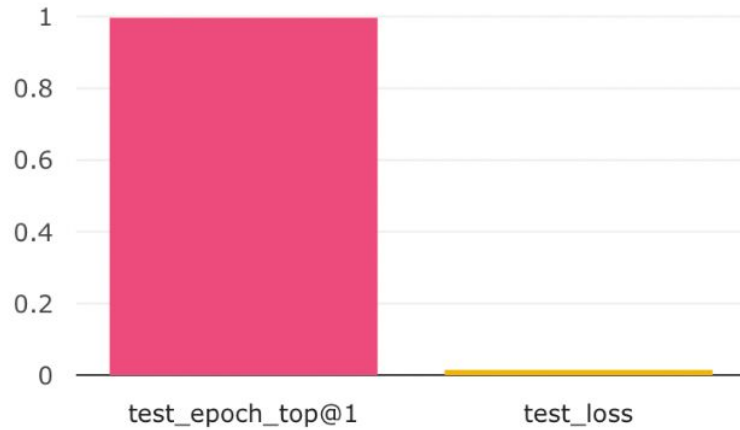


train\_loss

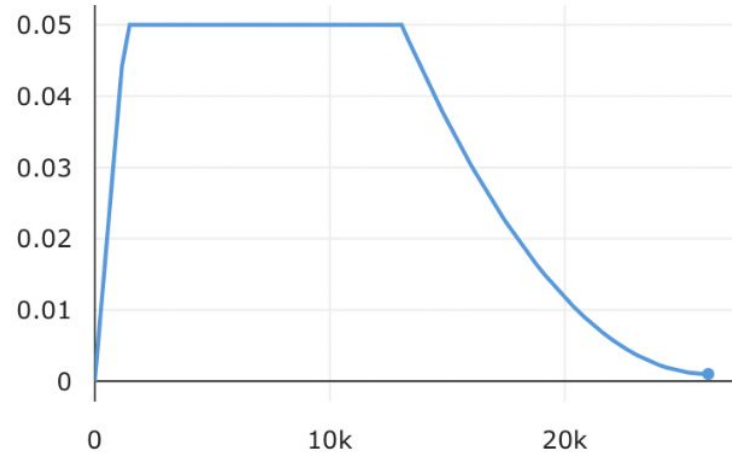


# Эксперимент 6

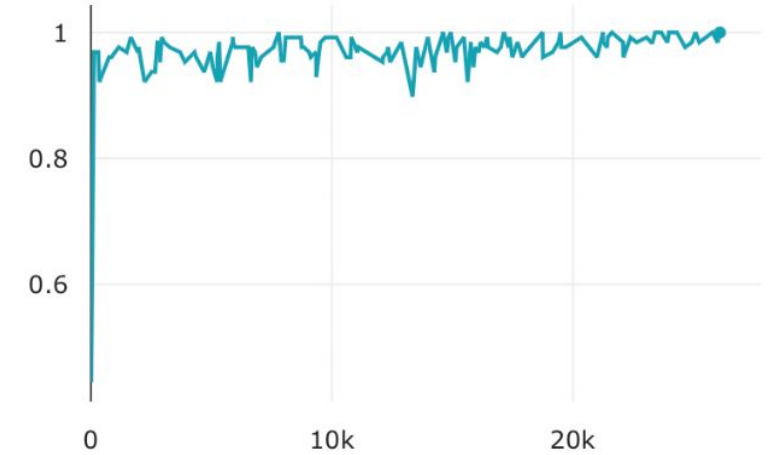
test\_epoch\_top@1, test\_loss



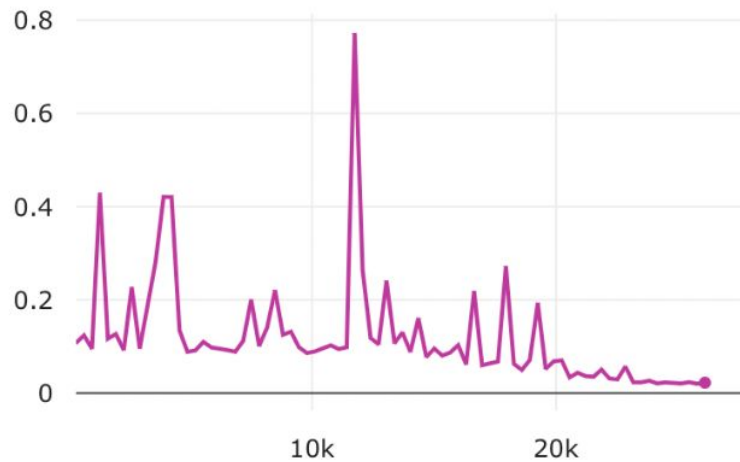
learning\_rate



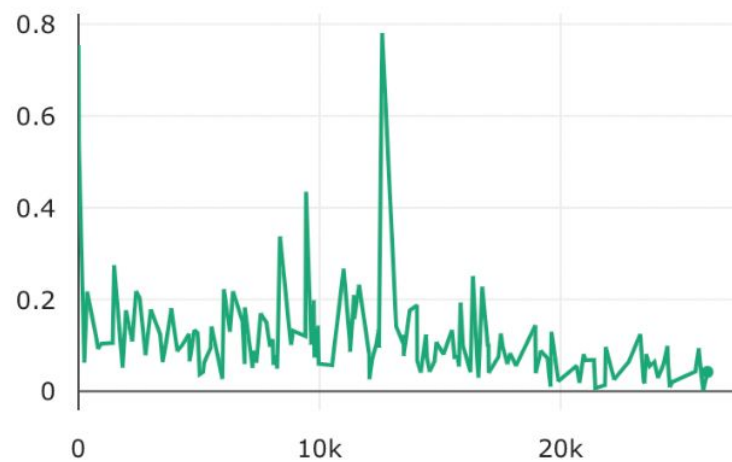
training\_batch\_accuracy\_top\_1



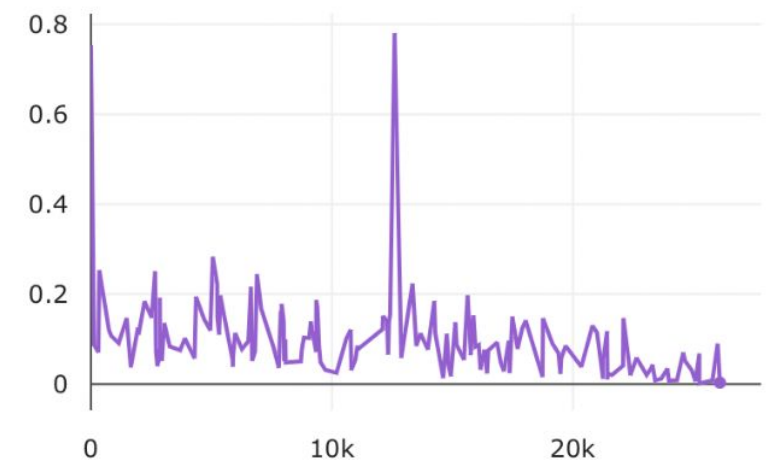
val\_loss



loss



train\_loss

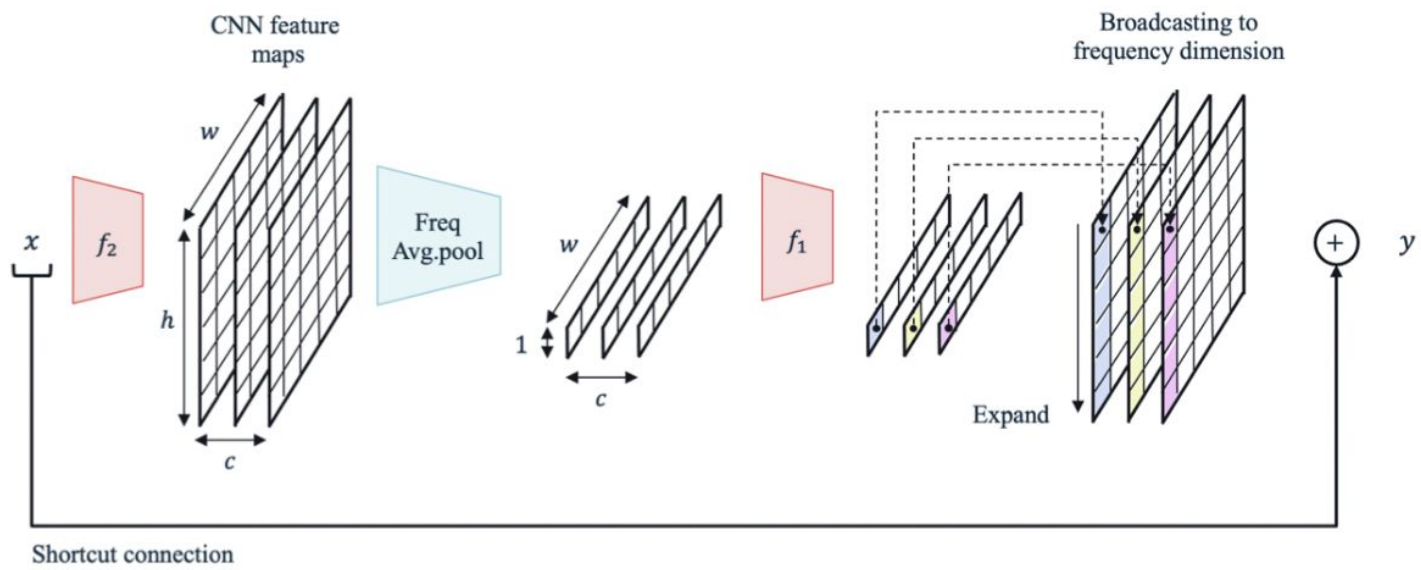


09

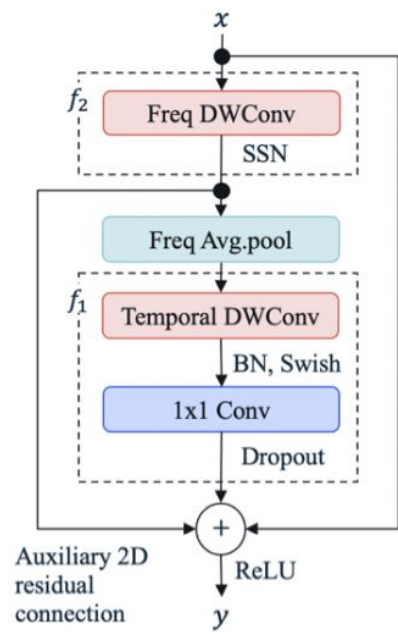
## Задумка НС

ОСНОВА - **Broadcasted Residual Learning**

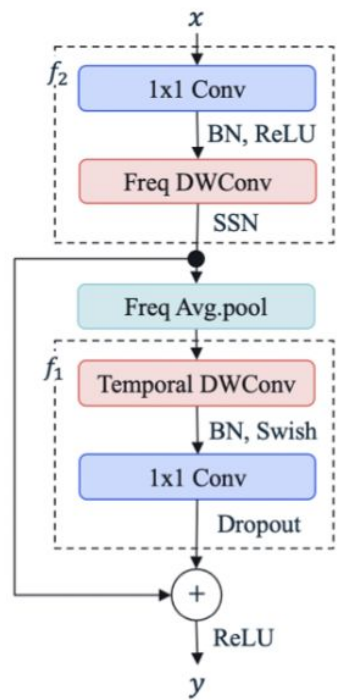
Broadcasted Residual Learning



Normal block



Transition block





Спасибо за  
внимание!

