

# Методы аугментации аудиоданных

Лукьянов Павел Александрович

Научный руководитель:

д.ф-м.н., профессор

Дьяконов Александр Геннадьевич

Москва, 2022

# Аугментация аудиоданных

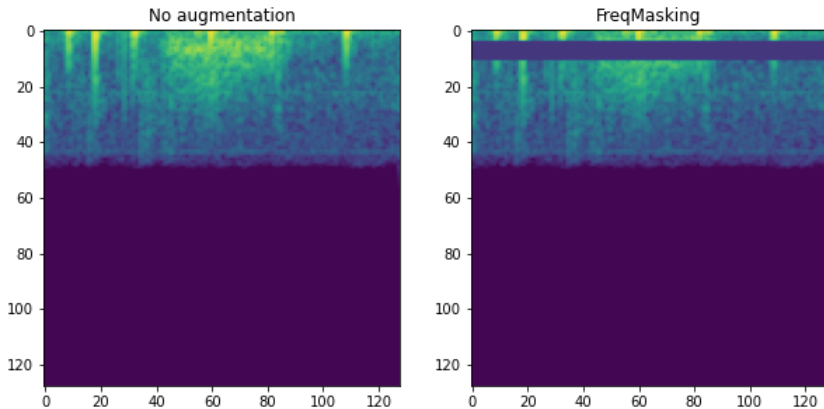


Рис.: FreqMasking

# SwapVerticalStripes

$T$  — параметр метода.

В результате применения аугментации:

- 1  $t \sim U\{0, T\}$ ,  $t_1 \sim U\{t, \text{TimeSize} - 1 - t\}$ ,  
 $t_2 \sim U\{t, \text{TimeSize} - 1 - t\}$ ,  $|t_1 - t_2| \geq t$ .
- 2  $S[:, t_1 : t_1 + t - 1] \leftrightarrow S[:, t_2 : t_2 + t - 1]$ .

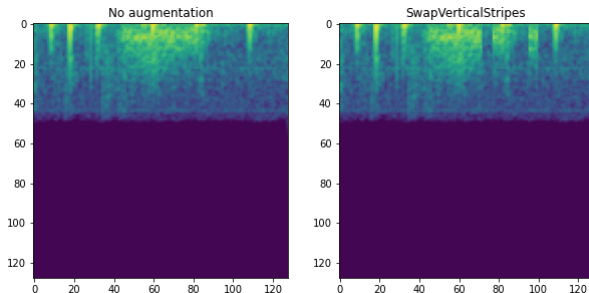


Рис.: SwapVerticalStripes

# SwapNeighboringStripes

$T$  — параметр метода.

В результате применения аугментации:

- 1  $t \sim U\{0, T\}, t_0 \sim U\{t, \text{TimeSize} - 1 - t\}.$
- 2  $S[:, t_0 : t_0 + t - 1] \leftrightarrow S[:, t_0 - t : t_0 - 1].$

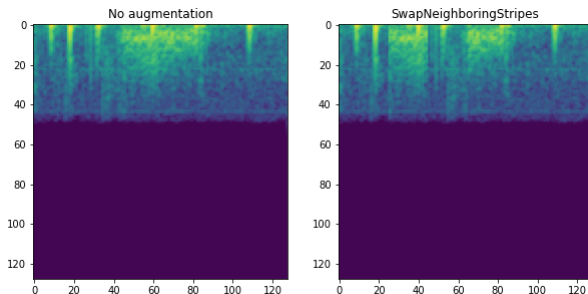


Рис.: SwapNeighboringStripes

# SwapSeveralStripes

$T, N$  — параметры метода,  $n \sim U\{0, N\}$ .

В результате применения аугментации (процедура повторяется  $n$  раз):

- 1  $T_0 = \lfloor \frac{T}{n} \rfloor$ ,  $t \sim U\{0, T_0\}$ ,  $t_1 \sim U\{t, \text{TimeSize} - 1 - t\}$ ,  
 $t_2 \sim U\{t, \text{TimeSize} - 1 - t\}$ ,  $|t_1 - t_2| \geq t$ .
- 2  $S[:, t_1 : t_1 + t - 1] \leftrightarrow S[:, t_2 : t_2 + t - 1]$ .

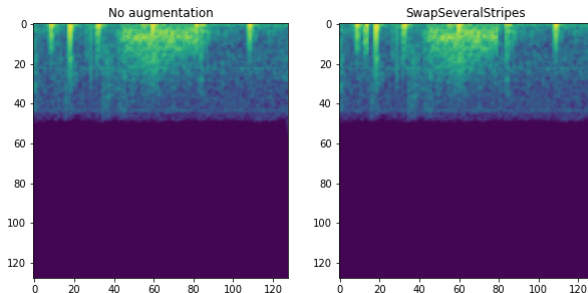


Рис.: SwapSeveralStripes

Метод аугментации	resnet18	resnet50
Аугментация отсутствует	81.98 $\pm$ 2.34	82.23 $\pm$ 2.4
SwapVerticalStripes	<b>83.2 <math>\pm</math> 1.3</b>	<b>83.65 <math>\pm</math> 1.07</b>
SwapNeighboringStripes	81.62 $\pm$ 0.69	<b>83.4 <math>\pm</math> 1.71</b>
SwapSeveralStripes	<b>83.55 <math>\pm</math> 0.49</b>	<b>84.42 <math>\pm</math> 1.92</b>

**Таблица:** Результаты экспериментов (Heartbeat Sounds) с предлагаемыми методами аугментации SwapVerticalStripes, SwapNeighboringStripes, SwapSeveralStripes. Метрика качества — процент верно классифицированных объектов.

Метод аугментации	resnet18	resnet50
Аугментация отсутствует	$0.66 \pm 0.034$	$0.692 \pm 0.04$
SwapVerticalStripes	<b><math>0.699 \pm 0.029</math></b>	$0.681 \pm 0.038$
SwapNeighboringStripes	<b><math>0.69 \pm 0.029</math></b>	$0.7 \pm 0.027$
SwapSeveralStripes	<b><math>0.687 \pm 0.026</math></b>	<b><math>0.709 \pm 0.029</math></b>

**Таблица:** Результаты экспериментов (Heartbeat Sounds) с предлагаемыми методами аугментации SwapVerticalStripes, SwapNeighboringStripes, SwapSeveralStripes. Метрика качества — сбалансированная точность.

Метод аугментации	resnet18	resnet50
Аугментация отсутствует	74.3 $\pm$ 3.03	73.0 $\pm$ 3.24
SwapVerticalStripes	<b>76.6 <math>\pm</math> 2.67</b>	<b>75.6 <math>\pm</math> 3.68</b>
SwapNeighboringStripes	<b>75.6 <math>\pm</math> 2.75</b>	71.4 $\pm$ 4.91
SwapSeveralStripes	<b>75.4 <math>\pm</math> 2.18</b>	72.7 $\pm$ 3.4

**Таблица:** Результаты экспериментов (GTZAN) с предлагаемыми методами аугментации SwapVerticalStripes, SwapNeighboringStripes, SwapSeveralStripes. Метрика качества — процент верно классифицированных объектов.



# Алгоритм применения методов аугментации

---

## Algorithm 1 Предлагаемый алгоритм

---

$\text{Augmentations} = \{\text{Augment}_1, \text{Augment}_2, \dots, \text{Augment}_n\}$  — заданный набор аугментаций,

$\text{Augment}$  — случайно выбранная аугментация из  $\text{Augmentations}$ ,

$(X_{\text{val}}, y_{\text{val}})$  — валидационный датасет,

$(X_{\text{train}}, y_{\text{train}})$  — обучающая выборка,

$f$  — метрика качества,

$M$  — число эпох обучения нейронной сети.

**Цикл от  $j = 1$  до  $M$  выполнять**

    train-шаг с применением  $\text{Augment}$ ,

    вычисление  $F_i = f(\text{Augment}_i(X_{\text{val}}), y_{\text{val}})$ ,  $i = \overline{1, n}$ ,

$\text{Augment} = \text{Augment}_k$ , где  $k = \text{argmin}_k(F_k)$ .

**Конец цикла**

---

Метод аугментации	resnet18	resnet50
Аугментация отсутствует	$81.98 \pm 2.34$	$82.23 \pm 2.4$
RandAugment	$83.1 \pm 0.92$	$84.57 \pm 1.3$
Предлагаемый алгоритм	<b><math>86.65 \pm 0.67</math></b>	<b><math>86.75 \pm 0.76</math></b>

**Таблица:** Результаты экспериментов (Heartbeat Sounds) с предлагаемым алгоритмом применения методов аугментации. Метрика качества — процент верно классифицированных объектов.

Метод аугментации	resnet18	resnet50
Аугментация отсутствует	$0.66 \pm 0.034$	$0.692 \pm 0.04$
RandAugment	$0.713 \pm 0.031$	$0.677 \pm 0.036$
Предлагаемый алгоритм	<b><math>0.762 \pm 0.023</math></b>	<b><math>0.753 \pm 0.02</math></b>

**Таблица:** Результаты экспериментов (Heartbeat Sounds) с предлагаемым алгоритмом применения методов аугментации. Метрика качества — сбалансированная точность.

Метод аугментации	resnet18	resnet50
Аугментация отсутствует	$74.3 \pm 3.03$	$73.0 \pm 3.24$
RandAugment	$75.0 \pm 2.61$	$74.9 \pm 2.63$
Предлагаемый алгоритм	<b><math>76.8 \pm 1.75</math></b>	$72.2 \pm 2.8$

**Таблица:** Результаты экспериментов (GTZAN) с предлагаемым алгоритмом применения методов аугментации. Метрика качества — процент верно классифицированных объектов.

В процессе выполнения работы получены следующие результаты:

- предложен и реализован метод аугментации аудиоданных `SwapVerticalStripes`, основанный на перестановке вертикальных полос в мел-спектрограмме, а также его модификации `SwapNeighboringStripes`, `SwapSeveralStripes`,
- проведены вычислительные эксперименты, показывающие возможную применимость предложенного метода `SwapVerticalStripes` и его модификаций в задаче аудиоклассификации,

# Заключение

- предложен и реализован алгоритм применения методов аугментации аудиоданных с выбором конкретного метода аугментации после каждой эпохи обучения,
- проведены вычислительные эксперименты, показывающие существенное преимущество предложенного алгоритма над алгоритмом RandAugment в задаче аудиоклассификации Heartbeat Sounds Classification.

По результатам работы сделан доклад на международной научной конференции студентов, аспирантов и молодых ученых «Ломоносов-2022».