

Методы аугментации аудиоданных

Лукиянов Павел Александрович

Научный руководитель:

д.ф-м.н., профессор

Дьяконов Александр Геннадьевич

Москва, 2021

Аугментация изображений

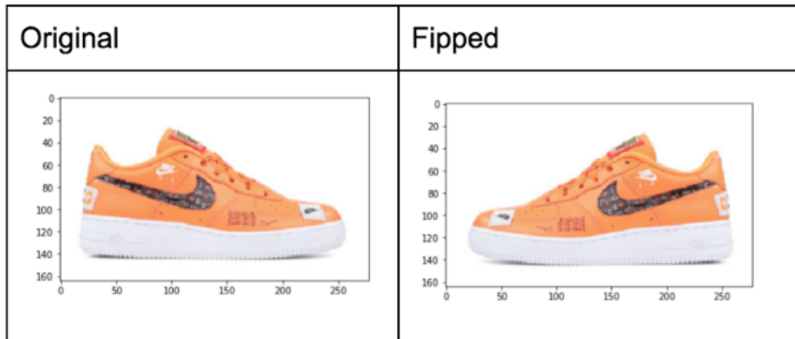


Figure: Flipping

TimeCycleShift

Циклический сдвиг всех значений мел-спектрограммы относительно временной оси влево или вправо на $|\text{shift}|$, где $\text{shift} \sim U\{-\text{max_shift}, \text{max_shift}\}$, max_shift - параметр аугментации.

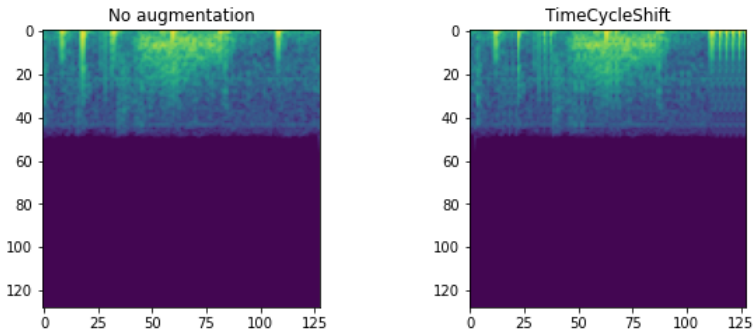


Figure: TimeCycleShift

TimeRandomSwap

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

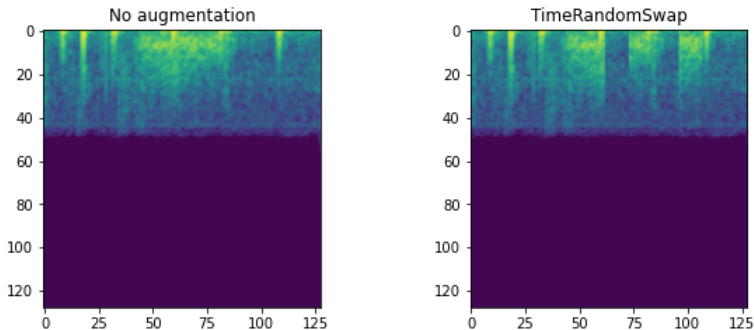


Figure: TimeRandomSwap

TimeSwapAugmentation

$t \sim U\{0, T\}$, $t_0 \sim U\{t, \text{TimeSize} - 1 - t\}$, T - параметр аугментации.

$$S[:, t_0 : t_0 + t - 1] \leftrightarrow S[:, t_0 - t : t_0 - 1]$$

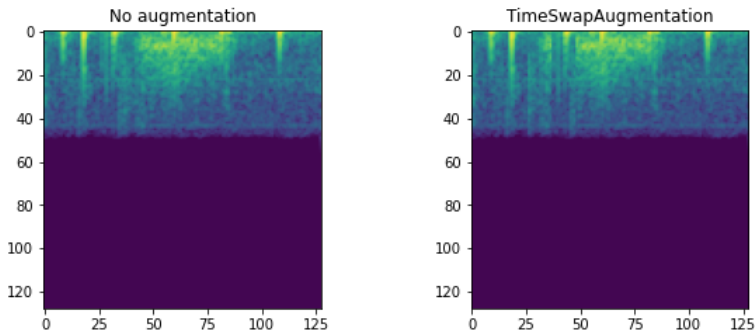


Figure: TimeSwapAugmentation

Метод аугментации	R18 + AM	R50 + AM
No Augmentation	0.954 \pm 0.01	0.953 \pm 0.009
TimeMasking	0.952 \pm 0.006	0.956 \pm 0.006
FreqMasking	0.952 \pm 0.004	0.957 \pm 0.004
Noise	0.958 \pm 0.006	0.951 \pm 0.009
RandomErasing	0.962 \pm 0.005	0.951 \pm 0.01
TimeShift	0.961 \pm 0.006	0.957 \pm 0.003
TimeCycleShift	0.962 \pm 0.006	0.956 \pm 0.014
TimeSwapAugmentation	0.957 \pm 0.009	0.95 \pm 0.011
TimeRandomSwap	0.96 \pm 0.003	0.96 \pm 0.005

Table: Результаты экспериментов (AudioMnist)

Метод аугментации	R18 + HB	R50 + HB
No Augmentation	0.83 \pm 0.016	0.824 \pm 0.018
TimeMasking	0.829 \pm 0.014	0.826 \pm 0.007
FreqMasking	0.829 \pm 0.014	0.825 \pm 0.013
Noise	0.837 \pm 0.009	0.821 \pm 0.015
RandomErasing	0.823 \pm 0.01	0.817 \pm 0.013
TimeShift	0.866 \pm 0.014	0.863 \pm 0.015
TimeCycleShift	0.87 \pm 0.01	0.872 \pm 0.017
TimeSwapAugmentation	0.835 \pm 0.015	0.833 \pm 0.015
TimeRandomSwap	0.835 \pm 0.008	0.823 \pm 0.021

Table: Результаты экспериментов (HeartBeatSounds)

Алгоритм применения методов аугментации

$Augmentations = \{Augment_1, Augment_2, \dots, Augment_n\}$ - заданный набор аугментаций.

(X_{val}, y_{val}) - валидационный датасет, f - метрика качества

$Augmentation(X_{val}) = \{Augmentation(x) \mid \forall x \in X_{val}\}$

$Augmentation_0$ - случайно выбранная аугментация из $Augmentations$

На каждой итерации:

- train-шаг с применением $Augmentation_0$
- $F = [f(Augment_1(X_{val}), y_{val}), \dots, f(Augment_n(X_{val}), y_{val})]$
- $Augmentation_0 = Augment_i$, где $i = \operatorname{argmin}(F)$

Метод аугментации	R18 + AM	R50 + AM
No Augmentation	0.951 ± 0.012	0.946 ± 0.009
Random choice	0.952 ± 0.013	0.946 ± 0.013
Choice after each epoch	0.962 ± 0.004	0.956 ± 0.008

Table: TimeCycleShift, Noise, TimeRandomSwap, FreqMasking

Метод аугментации	R18 + AM	R50 + AM
No Augmentation	0.951 ± 0.012	0.946 ± 0.009
Random choice	0.96 ± 0.005	0.946 ± 0.009
Choice after each epoch	0.966 ± 0.004	0.95 ± 0.006

Table: TimeShift, Noise, TimeMasking, FreqMasking

Метод аугментации	R18 + HB	R50 + HB
No Augmentation	0.813 ± 0.008	0.818 ± 0.024
Random choice	0.83 ± 0.014	0.845 ± 0.016
Choice after each epoch	0.851 ± 0.024	0.850 ± 0.018

Table: TimeCycleShift, Noise, TimeRandomSwap, FreqMasking

Метод аугментации	R18 + HB	R50 + HB
No Augmentation	0.813 ± 0.008	0.818 ± 0.024
Random choice	0.84 ± 0.006	0.843 ± 0.025
Choice after each epoch	0.873 ± 0.016	0.858 ± 0.009

Table: TimeShift, Noise, TimeMasking, FreqMasking

Заключение

В процессе выполнения работы получены следующие результаты:

- Предложены и реализованы подходы к аугментации аудиоданных: TimeCycleShift, TimeSwapAugmentation, TimeRandomSwap
- Проведены вычислительные эксперименты, показывающие эффективность этих подходов
- Предложен и реализован алгоритм применения методов аугментации с выбором метода аугментации после каждой эпохи
- Проведены вычислительные эксперименты, показывающие возможную применимость этого алгоритма в задаче аудиоклассификации

Дальнейшая работа

- Проведение экспериментов на других датасетах, других задачах
- Исследование mixup-подходов
- Исследование различных стратегий маскирования