

Конференция Ломоносов-2022

Методы аугментации аудиоданных

Лукьянов Павел Александрович

Научный руководитель:

д.ф-м.н., профессор

Дьяконов Александр Геннадьевич

Москва, 2022

Аугментация аудиоданных

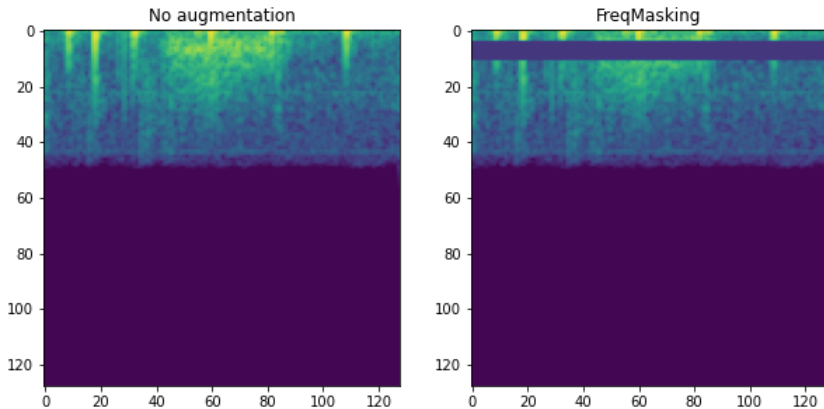


Рис.: FreqMasking

SwapVerticalStripes

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

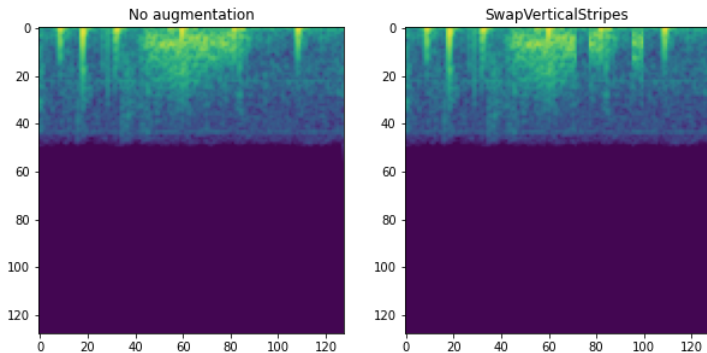


Рис.: SwapVerticalStripes

Метод аугментации	resnet18	resnet50
Аугментация отсутствует	81.98 ± 2.34	82.23 ± 2.4
SwapVerticalStripes	83.2 ± 1.3	83.65 ± 1.07

Таблица: Результаты экспериментов (Heartbeat Sounds) с предлагаемым методом аугментации SwapVerticalStripes

Метод аугментации	resnet18	resnet50
Аугментация отсутствует	74.3 ± 3.03	73.0 ± 3.24
SwapVerticalStripes	76.6 ± 2.67	75.6 ± 3.68

Таблица: Результаты экспериментов (GTZAN) с предлагаемым методом аугментации SwapVerticalStripes

Алгоритм применения методов аугментации

Algorithm 1 Предлагаемый алгоритм

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

Augment — случайно выбранная аугментация из 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 $j = 0$ до M выполнять

 train-шаг с применением Augment

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

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

Конец цикла

Эксперименты

Метод аугментации	resnet18	resnet50
Аугментация отсутствует	81.98 ± 2.34	82.23 ± 2.4
RandAugment	83.1 ± 0.92	84.57 ± 1.3
Предлагаемый алгоритм	86.65 ± 0.67	86.75 ± 0.76

Таблица: Результаты экспериментов (Heartbeat Sounds) с предлагаемым алгоритмом применения методов аугментации

Метод аугментации	resnet18	resnet50
Аугментация отсутствует	74.3 ± 3.03	73.0 ± 3.24
RandAugment	75.0 ± 2.61	74.9 ± 2.63
Предлагаемый алгоритм	76.8 ± 1.75	72.2 ± 2.8

Таблица: Результаты экспериментов (GTZAN) с предлагаемым алгоритмом применения методов аугментации

Заключение

В процессе выполнения работы получены следующие результаты:

- Предложен и реализован метод аугментации аудиоданных SwapVerticalStripes
- Проведены вычислительные эксперименты, показавшие возможную применимость предложенного метода в задаче аудиоклассификации
- Предложен и реализован алгоритм применения методов аугментации аудиоданных с выбором конкретного метода аугментации после каждой эпохи
- Проведены вычислительные эксперименты, показавшие возможную применимость предложенного алгоритма в задаче аудиоклассификации

- *Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, Quoc V. Le. RandAugment: Practical automated data augmentation with a reduced search space // arXiv preprint arXiv:1909.13719. — 2019.*
- *Bentley, P. and Nordehn, G. and Coimbra, M. and Mannor, S. The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results. — 2011.*
<http://www.peterjbentley.com/heartchallenge/index.html>
- *G. Tzanetakis and P. Cook. Musical genre classification of audio signals. // IEEE Transactions on Speech and Audio Processing. — 2002.*

- *Daniel S. Park, William Chan, Yu Zhang, Chung-Cheng Chiu, Barret Zoph, Ekin D. Cubuk, Quoc V. Le.*
SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition // *arXiv preprint arXiv:1904.08779*. - 2019.
- *Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun.*
Deep Residual Learning for Image Recognition // In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.