

Обзор методов аугментации звука

SpecAugment

<https://arxiv.org/pdf/1904.08779.pdf>

В оригинальной статье используется в задаче ASR.

Применяется к лог мел спектрограммам.

Признаки, извлекаемые сеткой, из спектрограмм после аугментации должны быть устойчивы к деформациям во времени, частичной потере частотной информации и частичной потере малых сегментов речи.

Применяются следующие деформации (на иллюстрации примеры):

- 1) Time Warping
- 2) Frequency masking
- 3) Time masking

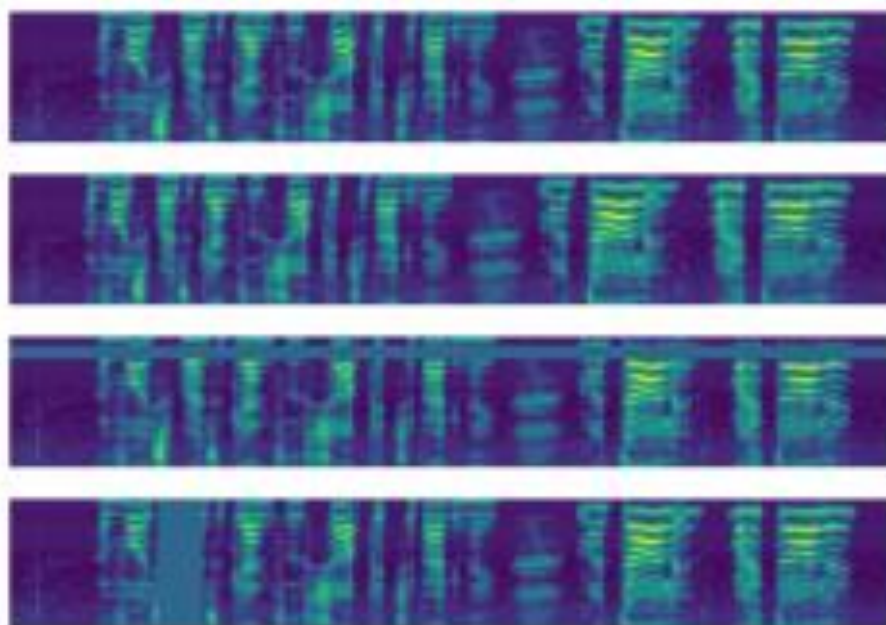


Figure 1: Augmentations applied to the base input, given at the top. From top to bottom, the figures depict the log mel spectrogram of the base input with no augmentation, time warp, frequency masking and time masking applied.

Multi-Window Data Augmentation

<https://arxiv.org/pdf/2010.09895.pdf>

Метод аугментации - использование оконных функций, применяемых к звуковому сигналу. В статье используется “Hamming” window.

https://en.wikipedia.org/wiki/Window_function

$$windowed_{signal}[t] = window[t] \times signal[t]$$

Размеры окна можно варьировать.

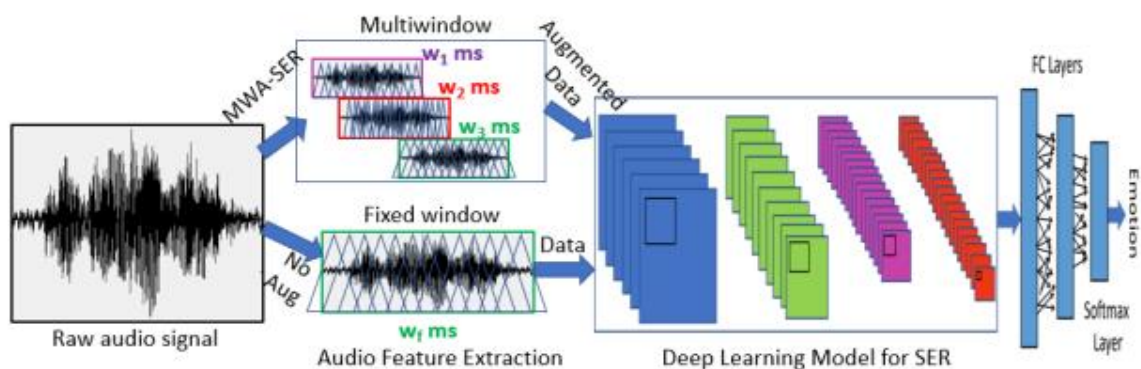


Схема представлена выше. В результате – улучшение есть, хотя сравнения с бейзлайнами мне показались немного странными.

Mask Detection and Breath Monitoring from Speech: on Data Augmentation, Feature Representation and Modeling

<https://arxiv.org/pdf/2008.05175.pdf>

Пример применения аугментации в задаче (Mask and Breathing Sub-Challenge).

Следующие аугментации использовались:

1) Speed perturbation (похоже на Time Warping)

Возмущение скорости применялось с коэффициентами 0.9, 1, 1.1.

Random Cropping и padding – чтобы не изменять размер изображения.

2) SpecAugment (без Time Wrapping)

3) Random erasing

Сводится к вырезанию случайных прямоугольников.

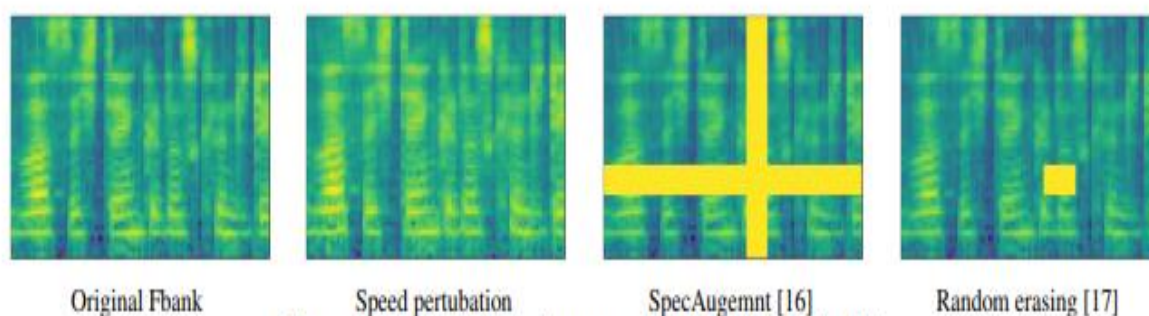


Figure 1: Effects of different data augmentation methods on the BSC

Surgical Mask Detection with Convolutional Neural Networks and Data Augmentations on Spectrograms

<https://arxiv.org/pdf/2008.04590.pdf>

Еще один пример использования аугментации звука в задаче Mask Detection.

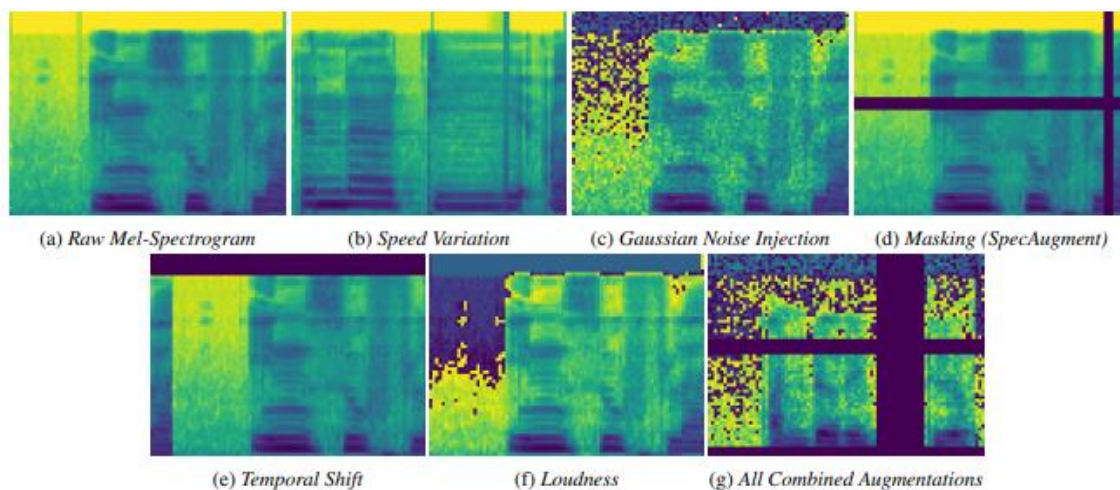


Figure 1: Examples for transformed mel-spectrograms, that are used in model training as given in section 3.

Библиотека для аугментации от Facebook:

<https://github.com/facebookresearch/WavAugment>

Есть также подходы к аугментации с использованием GAN'ов:

1. <https://arxiv.org/pdf/2004.05989.pdf>
2. <https://arxiv.org/pdf/2005.08447.pdf>
3. <https://arxiv.org/pdf/2006.10147.pdf>

Интересные статьи:

1. <https://arxiv.org/pdf/2008.12914.pdf>
2. <https://arxiv.org/pdf/2008.07231.pdf>
3. <https://arxiv.org/pdf/2008.03616.pdf>
4. <https://arxiv.org/pdf/2008.02480.pdf>
5. <https://arxiv.org/pdf/1910.13296.pdf>