RELATED WORK

В данной работе исследуются методы аугментации применительно к звуковым данным, а именно к мел-спектрограммам. Мел-спектрограммы подаются на вход нейронной сети как изображения, поэтому многие подходы аугментации изображений применимы и к аудиоданным. Например, метод Random Erasing [1], сводящийся к вырезанию случайных прямоугольников из изображения, может быть использован в задаче аудиоклассификации [2]. Также в задаче классификации звуковых данных применяются такие методы аугментации, как Shift Augmentation [3] - сдвиг мел-спектрограммы влево или вправо, Noise Augmentation [3] - добавление Гауссовского шума, Loudness Augmentation [3] - регулирование громкости, Speed augmentation [3] - ускорение или замедление аудиозаписи.

SpecAugment [4] - один из наиболее известных методов аугментации аудиоданных, который показал свою эффективность в задаче автоматического распознавания речи. Политика аугментации SpecAugment определяется 3 возможными преобразованиями: Time warping, Frequency masking, Time masking. В настоящее время известны некоторые модификации SpecAugment: SpliceOut [5], SpecAugment++ [6], FilterAugment [10].

Методы аугментации, основанные на mixup [7], также нашли применение в задачах, связанных с аудиоданными: MIXSPEECH [8], SPATIAL MIXUP [9]. Также возможны комбинации mixup [7] с другими методами аугментациями:
SpecMix [11], Cutmix [12].

Существуют подходы [13], [14] к аугментации звуковых данных, основанные на GAN [15]. Однако в данной работе подобные методы рассматриваться не будут.

Методы аугментации мел-спектрограмм не ограничиваются лишь преобразованиями соответствующих изображений. Существуют алгоритмы применения этих преобразованиях. В [16] предложена адаптивная весовая схема для аугментации временных рядов. Подобные алгоритмы применимы и для аугментации мел-спектрограмм.

Список литературы

[1] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, Yi Yang, "Random Erasing Data Augmentation". 2017.
https://arxiv.org/pdf/1708.04896.pdf

[2] Haiwei Wu, Lin Zhang, Lin Yang, Xuyang Wang, Junjie Wang, Dong Zhang, Ming Li, "Mask Detection and Breath Monitoring from Speech: on Data Augmentation, Feature Representation and Modeling". 2020.
https://arxiv.org/pdf/2008.05175.pdf

[3] Steffen Illium, Robert Muller, Andreas Sedlmeier and Claudia Linnhoff-Popien, "Surgical Mask Detection with Convolutional Neural Networks and Data Augmentations on Spectrograms". 2020.
https://arxiv.org/pdf/2008.04590.pdf

[4] 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". 2019.
https://arxiv.org/pdf/1904.08779.pdf

[5] Arjit Jain, Pranay Reddy Samala, Deepak Mittal, Preethi Jyoti, Maneesh Singh, "SpliceOut: A Simple and Efficient Audio Augmentation Method". 2021.
https://arxiv.org/pdf/2110.00046.pdf

[6] Helin Wang, Yuexian Zou, Wenwu Wang, "SpecAugment++: A Hidden Space Data Augmentation Method for Acoustic Scene Classification". 2021.
https://arxiv.org/pdf/2103.16858.pdf

[7] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, David Lopez-Paz, "mixup: Beyond Empirical Risk Minimization". 2017.
https://arxiv.org/pdf/1710.09412.pdf

[8] Linghui Meng, Jin Xu, Xu Tan, Jindong Wang, Tao Qin, Bo Xu, "MIXSPEECH: DATA AUGMENTATION FOR LOW-RESOURCE AUTOMATIC SPEECH

RECOGNITION". 2021.
https://arxiv.org/pdf/2102.12664.pdf

[9] Ricardo Falcon-Perez, Kazuki Shimada, Yuichiro Koyama, Shusuke Takahashi, Yuki Mitsufuji, "SPATIAL MIXUP: DIRECTIONAL LOUDNESS MODIFICATION AS DATA AUGMENTATION FOR SOUND EVENT LOCALIZATION AND DETECTION". 2021.
https://arxiv.org/pdf/2110.06126.pdf

[10] Hyeonuk Nam, Seong-Hu Kim, Yong-Hwa Park, "FILTERAUGMENT: AN ACOUSTIC ENVIRONMENTAL DATA AUGMENTATION METHOD". 2021.
https://arxiv.org/pdf/2110.03282.pdf

[11] Gwantae Kim, David K. Han, Hanseok Ko, "SpecMix : A Mixed Sample Data Augmentation method for Training withTime-Frequency Domain Features". 2021.
https://arxiv.org/pdf/2108.03020.pdf

[12] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, Youngjoon Yoo, "CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features". 2019.
https://arxiv.org/abs/1905.04899

[13] Nhat Truong Pham, Duc Ngoc Minh Dang, and Sy Dzung Nguyen, "Hybrid Data Augmentation and Deep Attention-based Dilated Convolutional-Recurrent Neural Networks for Speech Emotion Recognition". 2021.
https://arxiv.org/pdf/2109.09026.pdf

[14] Zengrui Jin, Mengzhe Geng, Xurong Xie, Jianwei Yu, Shansong Liu, Xunying Liu, Helen Meng, "Adversarial Data Augmentation for Disordered Speech Recognition". 2021.
https://arxiv.org/pdf/2108.00899.pdf

[15] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, "Generative Adversarial Nets". 2014.
https://arxiv.org/pdf/1406.2661.pdf

[16] Elizabeth Fons, Paula Dawson, Xiao-jun Zeng, John Keane, Alexandros Iosifidis, "Adaptive Weighting Scheme for Automatic Time-Series Data Augmentation". 2021. https://arxiv.org/pdf/2102.08310.pdf.