A proposal for AI project

"Data Augmentation for Text, Image, Sensor, and Audio Corpora"

Sristi Bhadani
Soham Joita

Supervised by
Prof. Dr.-Ing. Christian Bergler

Ostbayerische Technische Hochschule Amberg-Weiden
Department of Electrical Engineering, Media and Computer Science

July 1, 2024

Abstract

Data augmentation is a powerful technique to enhance the robustness and generalization capabilities of machine learning and deep learning models. It adds significant value in domains where obtaining large, diverse, and labeled datasets is difficult. The proposed research focuses on the development of data augmentation techniques across four data modalities: text, sensor, image, and audio. Each modality presents unique challenges and methods for augmentation, making it essential to tailor strategies to the specific characteristics of the data. The main goal of this study is to explore various augmentation techniques for text, sensor, image, and audio data, aiming to identify methods that enhance model performance, especially when training data is limited. This study also provides a comprehensive framework to implement augmentation strategies across different data types. For text data, techniques like synonym replacement, random insertion, and back-translation were explored. Sensor data augmentation included methods like jittering, scaling, and permutation. Image data augmentation involved approaches such as rotation, flipping, color adjustments, and blurring. For audio data, techniques such as time stretching, pitch shifting, and noise addition were applied. Tailoring augmentation techniques to the specific properties of text, sensor, image, and audio is important to achieve maximum effectiveness. These findings serve as a valuable reference for practitioners seeking to implement data augmentation in machine learning, highlighting the potential for performance gains. Overall, this study provides a comprehensive analysis of data augmentation techniques and offers insights into their application and impact across diverse data modalities.

1 Introduction and Motivation

In the field of artificial intelligence and deep learning, the quality and diversity of data play a crucial role in the success of models across domains. Image, audio, text, and sensor data are fundamental and cover a wide range of applications in machine learning and deep learning. Acquiring labeled data in sufficient quantity and diversity is challenging and often leads to models that overfit or do not generalize well. Data augmentation is a technique used in machine learning to artificially increase the size and diversity of a training dataset without collecting new data. The process helps create modified versions of existing data points to enhance the training process of models. For text data, methods like synonym replacement and back-translation are utilized to generate new textual variations. Sensor data augmentation includes techniques such as jittering and scaling to enhance the variability of sensor readings. Image data augmentation leverages transformations like rotation and color adjustments, while audio data augmentation uses methods such as noise addition and pitch shifting.

Despite the effectiveness of existing augmentation techniques, there is a noticeable gap in the literature regarding a unified framework that can integrate augmentation techniques across data modalities. Existing research often focuses on single data types, leaving a gap in understanding how different augmentation strategies perform when applied to text, sensor, image, and audio data.

The primary objectives of this study are to develop a unified augmentation library that can enhance audio, text, image, and sensor corpora by integrating state-of-the-art augmentation techniques, and to design a scalable and efficient framework for generating augmented data samples while preserving semantic coherence and integrity.

The proposed solution advances the field of machine learning, particularly in areas where data is scarce or expensive to obtain. This study aids in the development of more robust and generalizable models. By providing a unified solution for augmenting diverse data modalities, our framework can support further research in data augmentation, transfer learning, multi-modal training, and more.

This paper is organized as follows: Section 2 provides a comprehensive review of existing augmentation techniques across image, audio, text, and sensor data domains. Section 3 presents the data and its sources. Section 4 discusses the design and implementation of our unified data augmentation framework. In Section 5, we evaluate the performance of the framework on a variety of machine learning tasks. Finally, Section 6 discusses the implications of our findings and outlines potential avenues for future research.

2 Related Work

Recent advancements across various domains of machine learning have spurred innovative approaches in data augmentation, significantly enhancing model performance and robustness. In the realm of Natural Language Processing (NLP), techniques like back translation introduced by Sennrich et al.[6] (2016) have proven effective in improving neural machine translation using monolingual data. Wei and Zou [10] (2019) proposed straightforward yet powerful methods such as synonym replacement and random perturbations for text classification tasks, while Devlin et al.[3] (2019) showcased BERT's ability to generate high-quality augmented text, thereby enriching training datasets. Junczys-Dowmunt et al. [4](2018) applied machine translation for grammatical error correction, demonstrating the breadth of augmentation strategies in NLP.

In the research paper of "Sensor Data Augmentation by Resampling for Contrastive Learning for Human Activity Recognition," Wang et al. (2022) [9]detail various sensor data augmentation techniques. These include jittering, scaling, magnifying, rotating, inverting, reversing, permutation, time warping, cropping, and shuffling. They also introduce a novel resampling method, combining upsampling and downsampling, to simulate varying sensor frequencies and improve augmentation efficacy. These techniques enhance the robustness of contrastive learning frameworks for human activity recognition tasks.

Meanwhile, in computer vision, image data augmentation remains pivotal for enhancing deep learning model performance, especially crucial when labeled data is scarce. Beyond traditional geometric transformations and color space adjustments, recent advancements have seen the integration of Generative Adversarial Networks (GANs) and neural style transfer techniques. These methods generate highly realistic augmented images, significantly boosting model generalization across applications like medical imaging, where large datasets are often challenging to obtain. The comprehensive survey by Connor Shorten and Khoshgoftaar [2]("A Survey on Image Data Augmentation for Deep Learning") underscores the transformative impact of these techniques on computer vision tasks, highlighting their growing significance in advancing deep learning applications.

Furthermore, In recent years, audio augmentation has emerged as a pivotal technique to enhance the performance of deep learning models in various audio processing tasks, including speech recognition and acoustic scene classification. Comprehensive reviews have demonstrated the efficacy of methods like noise addition, pitch shifting, time stretching, and vocal tract length perturbation. Additionally, time-domain augmentation techniques like mixing and cropping have proven effective in acoustic scene classification tasks. These augmentation strategies are crucial for creating diverse and representative training datasets, ultimately leading to more robust and generalizable models. Notable contributions in this field include works and reviews by Ravi et al [5]. on comprehensive data augmentation techniques for speech and audio processing.

Together, these diverse advancements underscore the evolving landscape of data augmentation techniques, driving improvements in model robustness, generalization capabilities, and performance across various domains of machine learning and AI. By extending existing methodology our aim is to contribute towards the advancement of multi modal data augmentation techniques and help in development of more robust models.

3 Data Materials

This project requires four different types of datasets. The datasets are sourced from the Kaggle platform and zedge.net.

3.1 Text dataset

The name of this dataset is "Reviews,[8]" and it is in CSV format. This dataset contains approximately 500,000 reviews of Amazon fine foods. It includes product and user information, ratings, and plain text reviews. Additionally, it encompasses reviews from all other Amazon categories. We are considering Summary column from the dataset for our research.

3.2 Sensor dataset

For this project, we use live smoke sensor dataset [7], which is in CSV format. This data is collected from IoT devices deployed in various environments such as normal indoor, normal outdoor, indoor wood fire, firefighter training areas, indoor gas fire, outdoor wood, coal, and gas grill, and outdoor high humidity areas. The dataset contains 60,000 readings with UTC timestamps.

3.3 Image dataset

We conducted experiments using three different datasets to get variety of image data. Hence, it comprises of animal, human being, vegetable data which were visualised.

3.4 Audio dataset

We have used the ringtone [1] for this dataset, took sample of 3 audios to augment it.

4 Methodology

The methodology of the study will revolve around the development of a unified data augmentation framework for image, audio, text, and sensor corpora. We will design and implement augmentation strategies tailored to each data modality, ensuring the augmented data introduces variability and simulates real-world conditions.

In our study, we will employ a wide array of text augmentation techniques to enrich and diversify our dataset, thus bolstering the robustness and generalizability of our models. Key techniques will include synonym replacement using WordNet to add lexical variety while preserving semantic meaning, and noise addition to simulate typographical errors, training models to handle minor text perturbations. We will also use methods like random insertion, deletion, and swapping to increase text variability. Additionally, back translation and style transfer techniques, along with advanced generation methods using probabilistic and graph-based models, will provide diverse and contextually coherent text variations. The BERT augmentation technique and grammar correction will ensure contextual integrity and text accuracy, respectively.

For sensor data, we will employ a clipping method to ensure values remain within the original range, preventing unrealistic results from noise addition. Gaussian noise will be generated using a normal distribution with mean 0.0 and standard deviation 0.01, while uniform noise will use specified lower and upper bounds. Jittering will be achieved through Gaussian distribution. The GenerateRandomCurves function will add smooth, random curves via cubic splines, introducing realistic variations. The Permutation function will randomly permute data segments, enhancing diversity by splitting each row, shuffling segments, and reassembling them in a new order.

The augmentation function will dynamically apply a range of transformations based on the provided configuration for image data. The Compose method will sequentially combine multiple transformations, enabling complex augmentation pipelines. Resize will scale images while maintaining the aspect ratio. RandomHorizontalFlip and RandomVerticalFlip will introduce orientation variability by flipping images with a given probability. RandomRotation will enhance rotational robustness by rotating images by specified degrees. Grayscale will simplify color channels by converting images to grayscale. ColorJit-

ter will adjust brightness and hue, simulating different lighting conditions. RandomAffine will apply transformations like translation, scaling, and rotation, altering image geometry. GaussianBlur will mimic out-of-focus effects using a Gaussian kernel. RandomPosterize will reduce color bit-depth, creating varied intensity levels. CenterCrop will focus on the main subject by cropping the central region of an image. RandomPerspective will introduce perspective distortions, providing different viewpoints. Techniques like RandAugment, AugMix, and TrivialAugmentWide will automate augmentation policies, applying random and mixed transformations to enhance image variability and robustness. These methods will collectively ensure a wide range of transformations, aiding in training machine learning models that generalize well across different scenarios.

Several audio augmentation techniques will be applied to create diverse and representative datasets. The apply augmentation function will dynamically apply various techniques based on the provided configuration. The add noise method will introduce random noise controlled by a noise factor, simulating different environments. The change speed function will adjust playback speed, creating variations in tempo and rhythm. The reverse audio method will reverse the audio signal, providing a backward playback effect. The slow down audio function will reduce playback rate, making the audio sound more drawn out. The add echo method will introduce an echo effect with specified delay and decay parameters, creating a reverberation effect. The pitch shift function will modify the pitch by a given number of steps, altering perceived frequency. The time masking method will apply a mask over a portion of the audio, simulating missing data and enhancing robustness against dropout. The add sonic boom effect function will create a dramatic effect by adding a delayed, increased signal with decay, simulating a sudden, powerful sound impact. Each augmentation will enhance the variability and robustness of audio datasets for machine learning applications, ensuring realistic variations without compromising data integrity.

References

[1] "audio dataset". In: (). URL: https://www.zedge.net/ringtone/a895e626-2d4e-4706-996a-5b48852a4ac3.

[2] "Connor and Khoshgoftaar". In: (2022). URL: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0197-0.

[3] "Devlin". In: (2019). URL: https://aclanthology.org/N19-1423/.

[4] "Junczys-Dowmunt". In: (2018). URL: https://aclanthology.org/N18-1055/.

[5] "Ravi". In: (2022). URL: https://arxiv.org/abs/2211.05047.

[6] "Sennrich". In: (2016). URL: https://aclanthology.org/P16-1009/.

[7] "sensor dataset". In: (). URL: https://www.kaggle.com/datasets/deepcontractor/smoke-detection-dataset.

[8] "text dataset". In: (). URL: https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews?select=Reviews.csv.

[9] "Wang". In: (2022). URL: https://arxiv.org/pdf/2109.02054.

[10] "Wei and Zou". In: (2019). URL: %7Bhttps://aclanthology.org/D19-1670/%7D.