Windowing Effects on SVM and CNN Performance

Pranjal Malik (M24CSA021)

Introduction

Windowing functions play a crucial role in spectral analysis by reducing spectral leakage in Short-Time Fourier Transform (STFT)-based feature extraction. This report compares the classification performance of two machine learning approaches—Support Vector Machines (SVM) and Convolutional Neural Networks (CNN)—when trained on spectral features extracted using three different windowing functions:

- Hamming Window
- Rectangular Window
- Hann Window

The evaluation is based on accuracy, precision, recall, and F1-score for different sound categories in the UrbanSound8K dataset.

Performance Comparison

Window Function	**SVM Accuracy**	**CNN Accuracy**
Rectangular	51%	70.29%
Hamming	49%	67.20%
Hann	38%	67.83%

Key Observations

- **CNN outperforms SVM significantly**, achieving ~70% accuracy compared to a maximum of 51% in SVM.
- The best-performing window remains **Rectangular** for both models, but the gap is larger in SVM than in CNN.
- **Hann window**, which performed the worst in SVM (38%), showed improvement in CNN (67.83%), indicating CNN's robustness in handling smoothed spectral features.

CNN Performance Highlights

- CNN does not struggle as much with **Hann windowing** as SVM does, suggesting that deep learning models can adapt to feature smoothing better than traditional classifiers.
- Higher recall and precision were observed across all classes compared to SVM.
- The **Rectangular window remains the optimal choice** in CNN as well, but the impact of windowing is reduced compared to SVM.

CNN Architecture

The CNN model used for classification is **AudioCNN**, consisting of the following layers:

- **Convolutional Layers**
  - Conv2D (1→16), Conv2D (16→32), Conv2D (32→64), each followed by ReLU activation.
  - MaxPooling after each convolution layer to reduce spatial dimensions.

- **Adaptive Pooling & Fully Connected Layers**
  - AdaptiveAvgPool2D to generate a fixed-size feature vector.
  - Flattening followed by a fully connected layer (64→128→10) with ReLU activations and softmax output.

Advantages of CNN over SVM

- Feature Learning Unlike SVM, CNN learns hierarchical features directly from spectrograms, reducing dependence on manually selected features.
- Higher Generalization CNNs handle variations in audio signals better than SVM.
- Improved Handling of Windowing Effects Unlike SVM, CNNs are less affected by different windowing functions, as seen in the Hann window results.

Conclusion

- CNN provides **significant improvements** over SVM for the UrbanSound8K dataset.
- The **Rectangular window remains the best choice**, but CNN mitigates the negative effects of Hann windowing.
- Future improvements should explore **transformer-based architectures (like Wav2Vec2)** or hybrid feature extraction to further enhance performance.