**A Light Convolutional GRU-RNN Deep Feature Extractor for ASV Spoofing Detection**

1. **Key innovations:**

* The aim of this work is to develop a single anti-spoofing system which can be applied to effectively detect all the types of spoofing attacks considered in the ASVspoof 2019 Challenge: text-to-speech, voice conversion and replay based attacks.
* The paper proposes the use of a Light Convolutional Gated Recurrent Neural Network (LC-GRNN) as a deep feature extractor to robustly represent speech signals as utterance-level embeddings, which are later used by a back-end recognizer which performs the final genuine/spoofed classification.
* This novel architecture combines the ability of light convolutional layers for extracting discriminative features at frame level with the capacity of gated recurrent unit based RNNs for learning long-term dependencies of the subsequent deep features.
* This paper proposes a hybrid light convolutional neural network (LCNN) [8] plus recurrent neural network (RNN) architecture which combines the ability of the LC NNs for extracting discriminative features at frame level with the capacity of gated recurrent unit (GRU) based RNNs for learning long-term dependencies of the subsequent deep features.
* Replace the fully connected layers inside the recurrent cells with LCNN layers in order to: (1) extract discriminative features at frame level, (2) learn long-term dependencies, and (3) integrate the extraction of frame-level deep features and the utterance-level identity vector into a single network.
* For the back-end, evaluated three different classifiers: support vector machine (SVM), linear discriminant analysis (LDA), and its probabilistic version (PLDA). The objective of the classifier is to assign a score indicating whether the utterance is genuine or spoofed.

1. **Reported Performance metrics:**

**2.1 Results on ASVspoof 2015:**

* The proposed system with PLDA classifier achieves the best performance in the known and unknown attacks, outperforming other state-of-the-art systems such as the CQCC + GMM and LTSS + MLP.
* It outperforms other similar deep feature extractors of the literature such as the CNN + RNN and FBANK + Best RNN, as well as our previous system FBANK + CNN + RNN.

**2.2 Results on ASVspoof 2017:**

* Proposed system with PLDA and scoring normalization achieves the best performance.
* It outperforms other state-of-the-art single systems such as the SCMC + GMM, LCNN + GMM and CNN + RNN, which were presented to the ASVspoof 2017 Challenge.

**2.3 Results on ASVspoof 2019 LA:**

* The authors evaluated their proposed system **LC-GRNN** combined with three scoring methods: PLDA, LDA, SVM.

**Compared Systems**:

* Baselines: **CQCC + GMM** and **LFCC + GMM**
* Submitted: **LC-GRNN + PLDA**, **LC-GRNN + LDA**, **LC-GRNN + SVM**
* **All LC-GRNN variants outperform** the baseline systems on both the **development and evaluation sets**.
* LC-GRNN + LDA Performs **22.37% better than CQCC + GMM.**
* **34.38% better than LFCC + GMM on the evaluation set.**
* Although LDA slightly outperforms PLDA, the **EER difference is minimal (0.06%)**, showing both are comparably effective.

**2.4 Results on ASVspoof 2019 PA:**

* The proposed LC-GRNN system clearly outperforms the baseline systems on the development and evaluation sets independently of the scoring classifier.
* Primary/single system (LC-GRNN + PLDA with scoring normalization) achieves a relative 74.69% and 79.80% better performance than CQCC + GMM and LFCC + GMM on the evaluation set, respectively.

1. **Promising aspects of this approach:**

* The LC-GRNN is designed to detect various spoofing attacks, including text-to-speech (TTS), voice conversion (VC), and replay attacks, within a common framework.
* The architecture combines light convolutional layers for extracting discriminative features at the frame level with gated recurrent unit (GRU) based RNNs, which learn long-term dependencies of the deep features.
* By replacing fully connected layers with LCNN layers inside the recurrent cells, more discriminative features can be extracted at the frame level.

1. **Potential limitations and challenges:**

* The model is trained and tested on the ASVspoof 2019 LA dataset, which limits exposure to unknown spoof types. This indicates a potential lack of robustness against zero-day spoofing techniques not present in training.
* Evaluation is conducted only on ASVspoof 2019 LA. There is no testing on cross-dataset performance (e.g., PA, other datasets), which raises concerns about practical generalization to real-world spoofing.
* The LC-GRNN does not offer insights into why it labels an input as spoofed, which is typical for black-box deep models. This can be a limitation in critical applications where explainability is important (e.g., security, forensics).