12
04
2017
BOLOGNA
LABORATORIO DELLE ARTI

INTELLIGENZA ARTIFICIALE:
DALL'UNIVERSITÀ ALLE AZIENDE.
LA RIVOLUZIONE DEL DEEP LEARNING

E4 COMPUTER ENGINEERING
NVIDIA
in collaborazione con IBM

Analisi dei segnali audio basata su Deep Learning per applicazione di sicurezza

Alessandro Neri, Francesco Calabrò, Federica Battisti, Marco Carli, Federico Colangelo,

BOLOGNA , 12 aprile 2017

LEONARDO
INGENUITY AT YOUR SERVICE

ROMA TRE
UNIVERSITÀ DEGLI STUDI

RadioLabs

I sistemi di audiosorveglianza

I sistemi audio forniscono informazioni utili per la rivelazione di particolari eventi in casi in cui i sistemi video non riescono a rilevarli in maniera altrettanto affidabile.

In questa ottica un sistema di audio sorveglianza può essere utilizzato come **COMPLEMENTO** alla videosorveglianza.

Finalità del Progetto

A BREVE TERMINE

Sviluppo di un sistema **AUTOMATICO** di classificazione **AUDIO** basato su algoritmi di *signal processing* e *machine learning* con **complessita' ridotta** del sistema di ripresa audio e con **bassi consumi** energetici

A LUNGO TERMINE

INTEGRAZIONE AUDIO-VIDEO per realizzare un sistema di sorveglianza **multimediale** automatizzato

Criticità

1. Difficoltà di sorvegliare contemporaneamente **PIÙ SORGENTI** audio con un **NUMERO RIDOTTO** di operatori (diversamente dai video)

2. Necessità di gestire in **MODO UNITARIO** fenomeni caratterizzati da **DURATE DIFFERENTI** (colpi di pistola, urla e rotture di vetri)

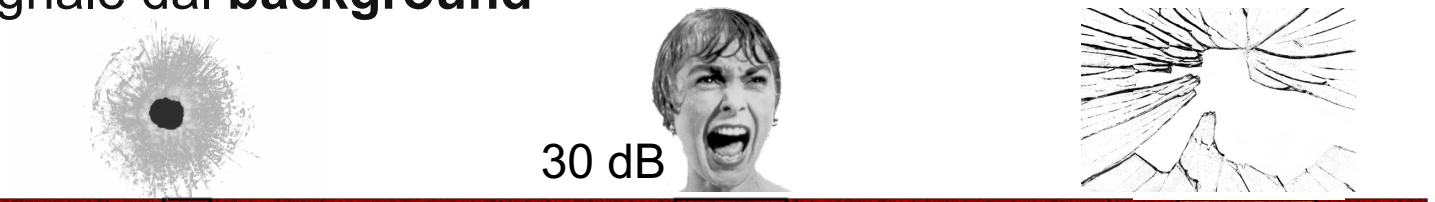
3. Necessità di reperire grandi quantità di **DATI REALISTICI** per la fase di **ADDESTRAMENTO** degli algoritmi di *machine learning*.

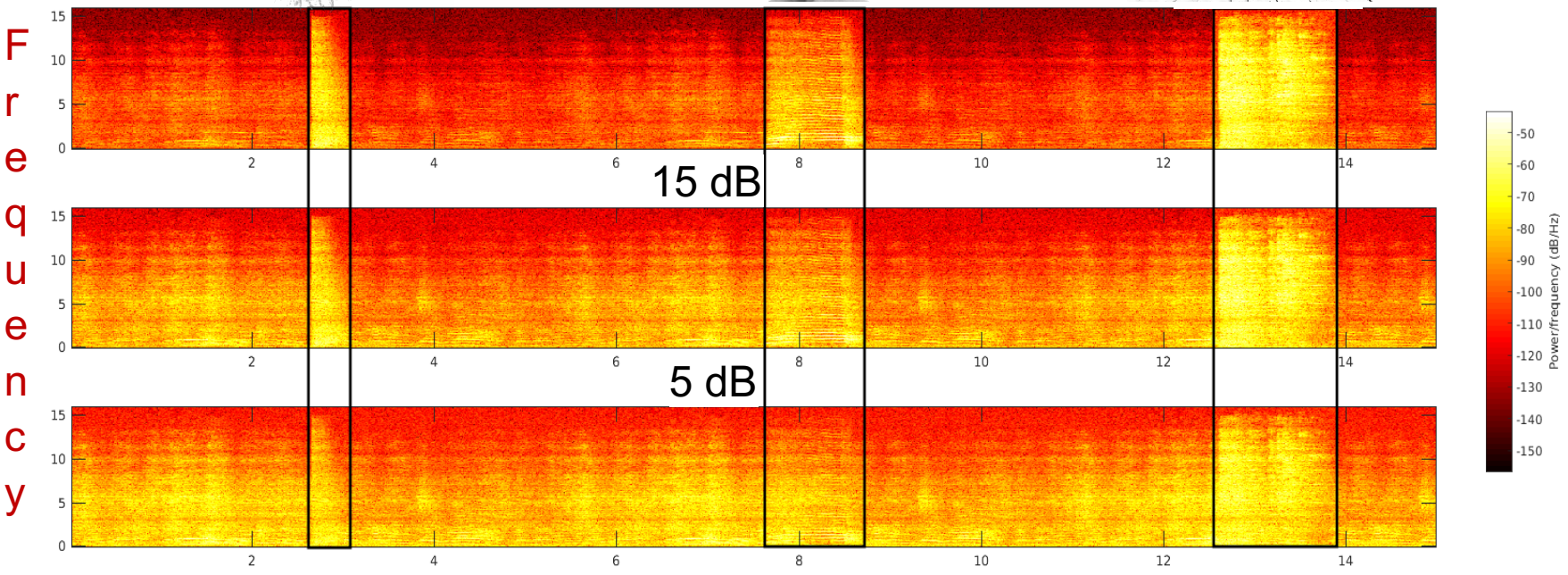
Eventi nel dominio della frequenza

- Ad **alti rapporti segnale rumore**, il problema principale è la discriminazione **interclasse**
- A **bassi rapporti segnale rumore**, bisogna anche discriminare il segnale dal **background**

Eventi nel dominio della frequenza

- Ad **alti rapporti segnale rumore**, il problema principale è la discriminazione **interclasse**
- A **bassi rapporti segnale rumore**, bisogna anche discriminare il segnale dal **background**

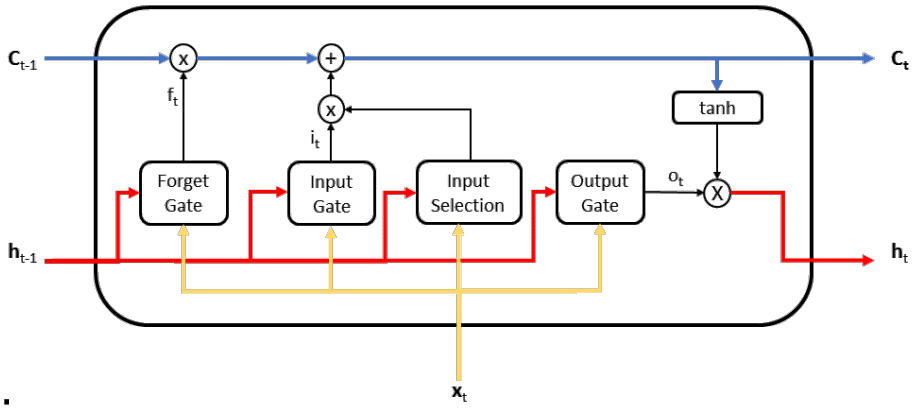




Recurrent Neural Nets

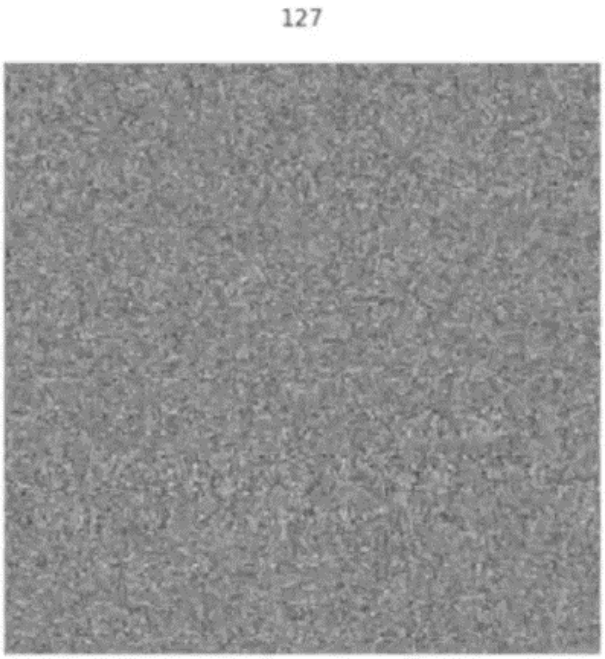
- Per comprendere il segnale audio è necessario tenere conto delle **dipendenze temporali**.
- Le **R**ecurrent **N**eural **N**et (**RNN**) nascono per questo scopo
  - Quelle di base non riescono a modellare dipendenze temporali a lungo termine

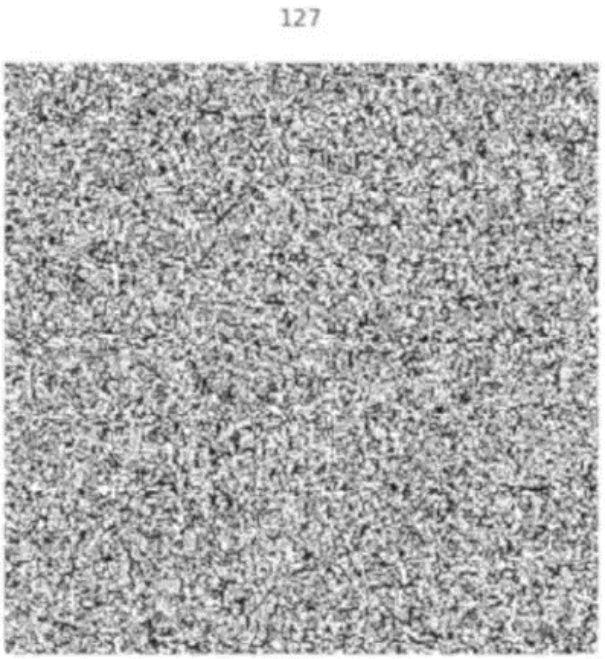
- Le RNN basate su **L**ong **S**hort **T**erm **M**emory (**LSTM**) sono in grado di apprendere rappresentazioni strutturate temporalmente, **selezionando** gli elementi chiave da ricordare e **dimenticando** quelli irrilevanti



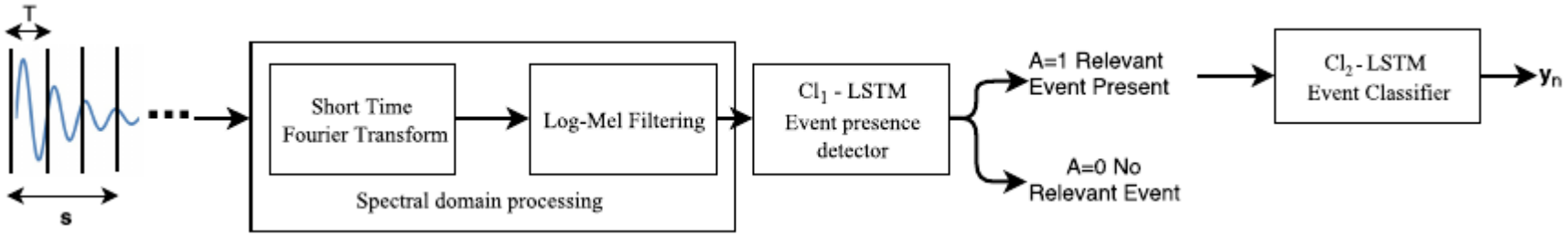
Gradienti: RNN classica vs RNN LSTM

- I gradienti sopravvivono molto più a lungo



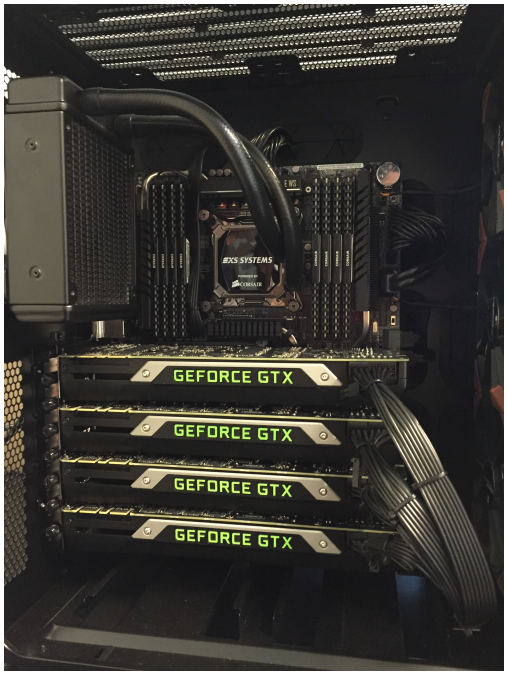


Architettura v. 1.0

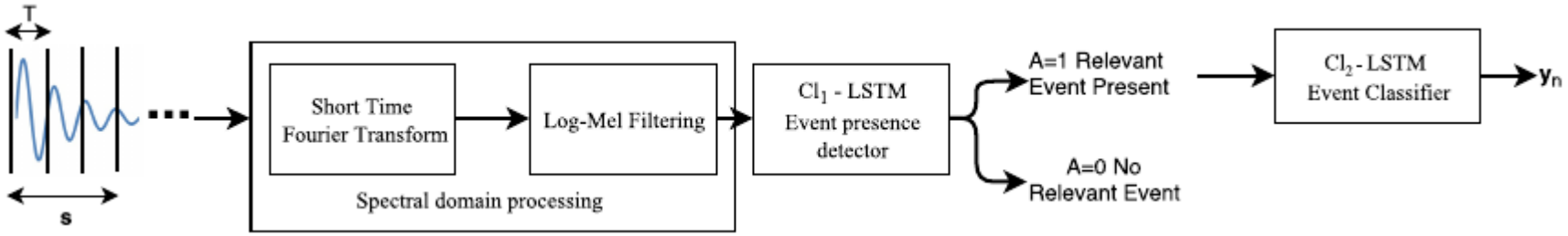


Rispetto allo stato dell'arte:

- 30 dB 90% → 99.9%

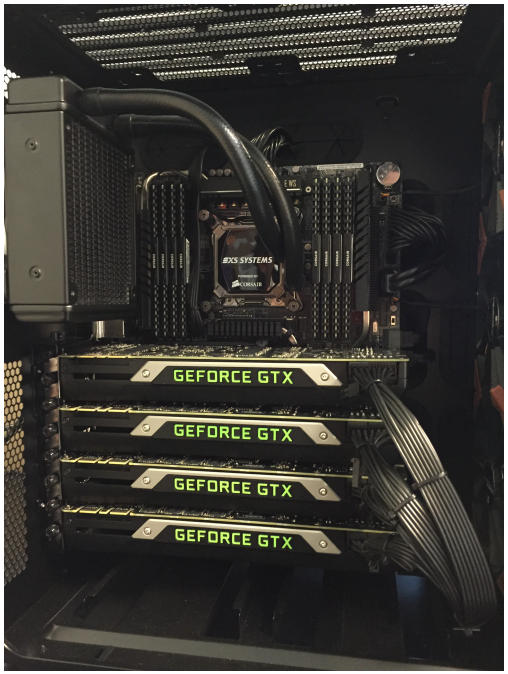


Architettura v. 1.0

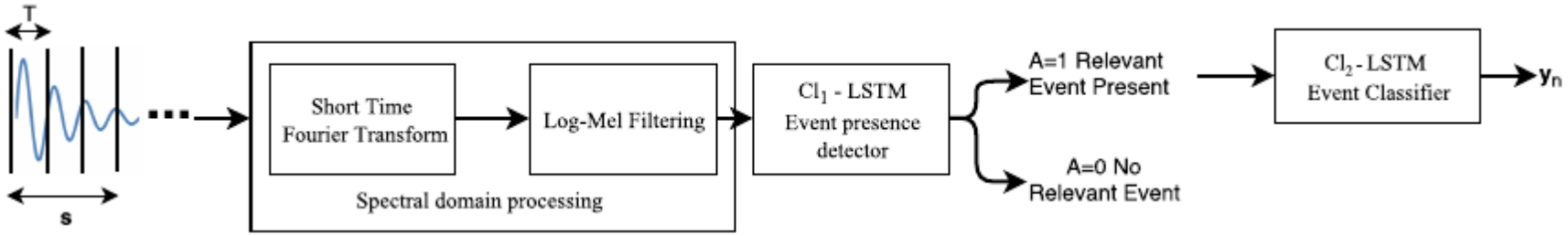


Rispetto allo stato dell'arte:

- 30 dB 90% → 99.9%
- 15 dB 87 % → 98.5%

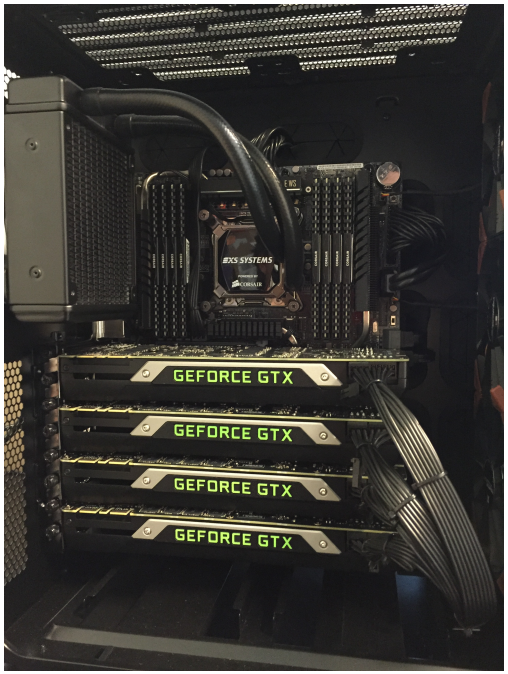


Architettura v. 1.0



Rispetto allo stato dell'arte:

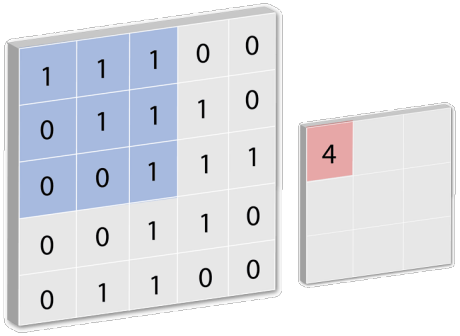
- 30 dB 90% → 99.9%
- 15 dB 87 % → 98.5%
- 5 dB 81.1% → 90.7%



Riconoscimento eventi 2.0

- I coefficenti Mel hanno buone prestazioni, ma perdono informazione…
  - … però rendono il problema trattabile
- Per migliorare la capacità di classificazione a bassi SNR serve
  - Maggiore risoluzione in frequenza
  - Ridurre la perdita di informazione

Uso di reti convoluzionali…

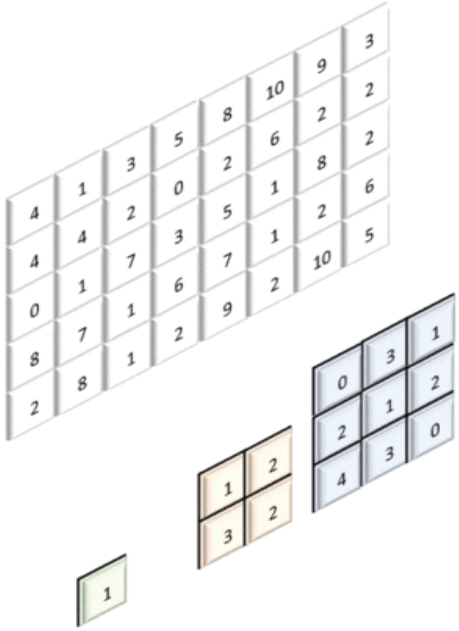


Riconoscimento eventi 2.0

- I coefficenti Mel hanno buone prestazioni, ma **perdono informazione**…
  - … però rendono il problema trattabile
- Per distinguere meglio a bassi SNR serve
  - **Risoluzione** in frequenza
  - Più informazione

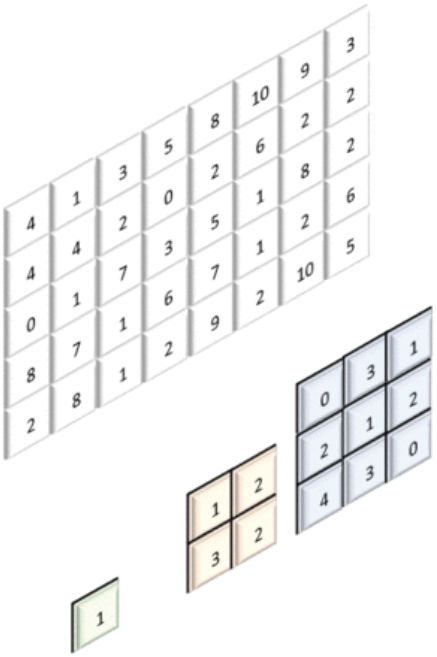
  Uso di reti convoluzionali…

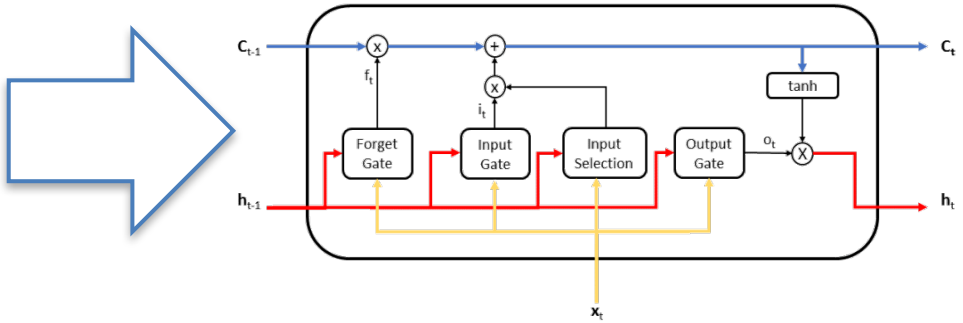
  …**multirisoluzione**



Combinazione di LSTM e CNN

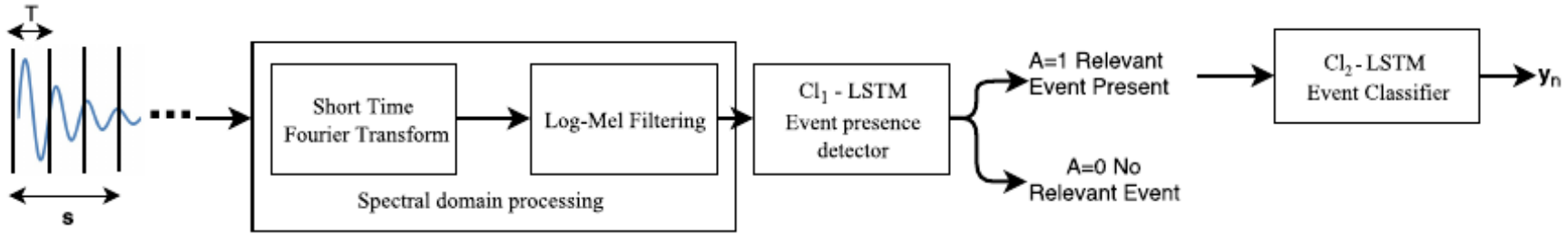
- Le CNN **estraggono l'informazione**
  - Resistenti al rumore, anche a bassi SNR
  - Possono lavorare su dati più grezzi
- Le celle LSTM apprendono le **dipendenze temporali**





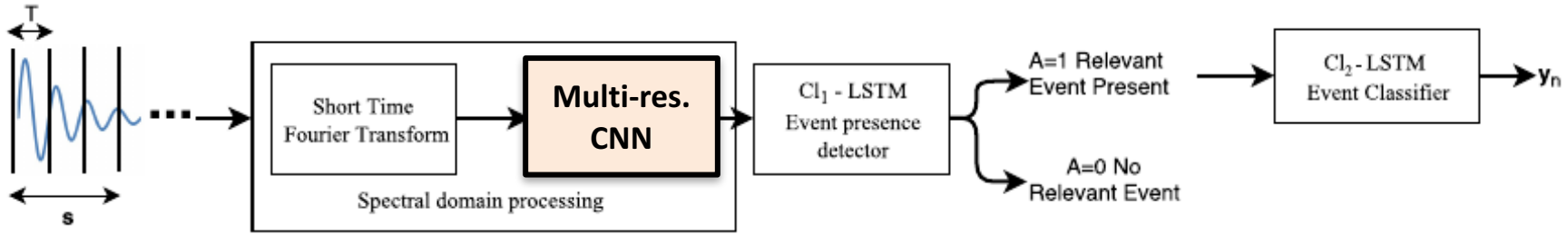
Combinazione di LSTM e CNN

- Le CNN **estraggono l'informazione**
  - Resistenti al rumore, anche a bassi SNR
  - Possono lavorare su dati più grezzi
- Le celle LSTM apprendono le **dipendenze temporali**



Combinazione di LSTM e CNN

- Le CNN **estraggono l'informazione**
  - Resistenti al rumore, anche a bassi SNR
  - Possono lavorare su dati più grezzi
- Le celle LSTM apprendono le **dipendenze temporali**



Prestazioni e attività future

- **Accuracy verificata al più basso livello di SNR (5dB)**
  - 81.1% → 90.7% → **94%**
- **Estensioni**
  - Rivelare la presenza di **SEGNALI STRUTTURATI** che possono mascherare gli eventi di interesse e mitigarne gli effetti
    - E.g. C'è qualcuno che canta?
  - Classificare gli eventi secondo una **CASISTICA** più ampia e **DETTAGLIATA**
    - E.g. Quale tipo di arma ha sparato?
  - Rendere i sensori più *smart*.

Attività in corso

Miglioramenti relativi ai dati di addestramento

- Impiego di un *dataset* acquisito sul campo per gli eventi di tipo *gunshot* al fine di **classificare il tipo di arma.**
- Registrazioni in situazioni eterogenee al fine di studiare/ottimizzare le prestazioni rispetto a **diversi tipi di ambienti**.

Grazie per l'attenzione

LEONARDO
INGENUITY AT YOUR SERVICE

ROMA TRE
UNIVERSITÀ DEGLI STUDI

RadioLabs