



# EXTRACCION DE RITMOS DE AUDIOS

Francisco José Durá Galiana

# Extracción de ritmos de audios

- Motivación y objetivos
- Procedimiento
- Metodología
- Resultados

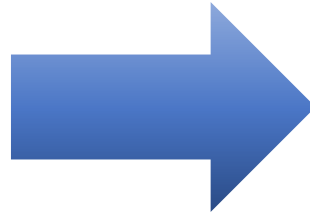
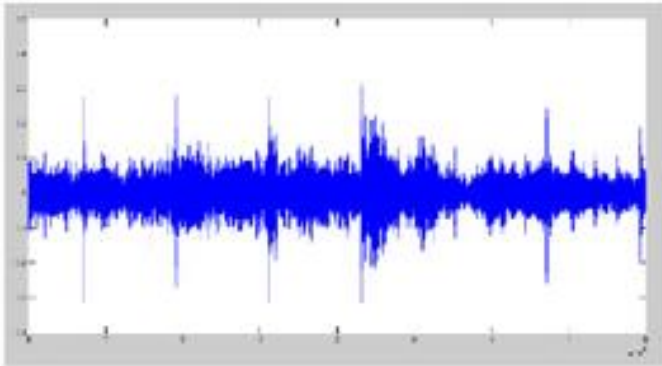
# Motivación y objetivos

# Motivación: ¿por qué predecir ritmos?

- Transcripción automática de audios
- Estudio (práctica) de música
- Inspiración/curiosidad

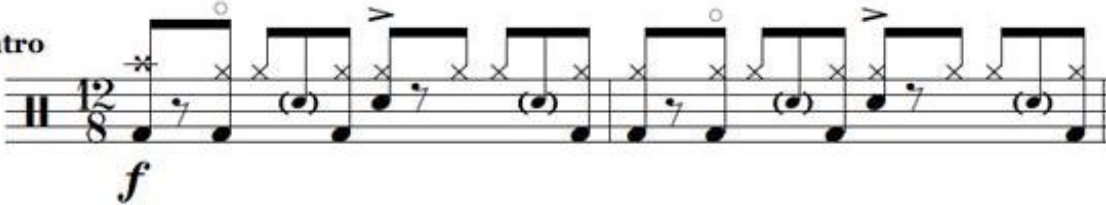


# Objetivo: Ideal



**Half-Time Shuffle** ♩ = 130

Intro



*f*

# Objetivo: Real

- Investigar posible algoritmo para extracción de ritmos
- Simplificación del problema: Género y tempo
  - Predecir Género
  - Predecir tempo
- Construir un predictor conjunto de ambas variables

# Procedimiento

# Procedimiento

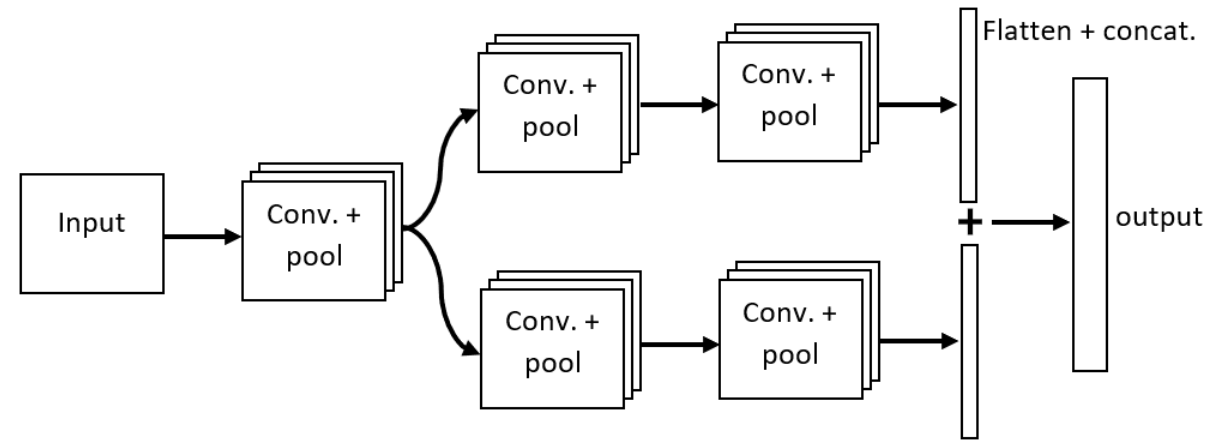
## 1. Red de predicción de género

1. Arquitectura simple
2. Arquitectura paralela

## 2. Red de predicción de tempo

1. Arquitectura simple
2. Arquitectura paralela

## 3. Red de predicción conjunta



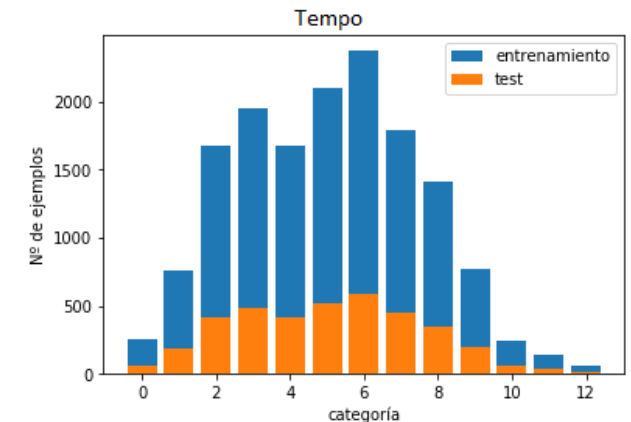
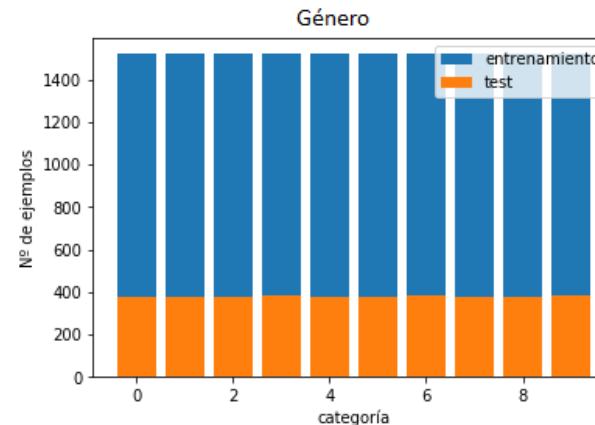
Ejemplo de arquitectura paralela



# Metodología

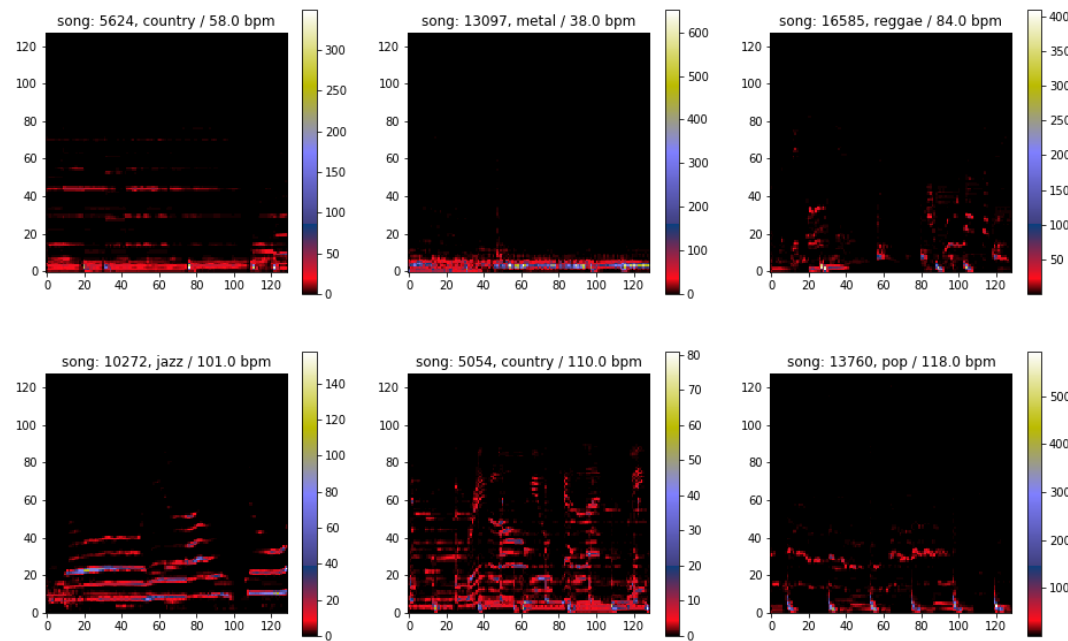
# Metodología I: Datos

- Búsqueda de bases de datos
- Análisis de variables objetivo
- Obtención de datos brutos
- Procesamiento de los datos

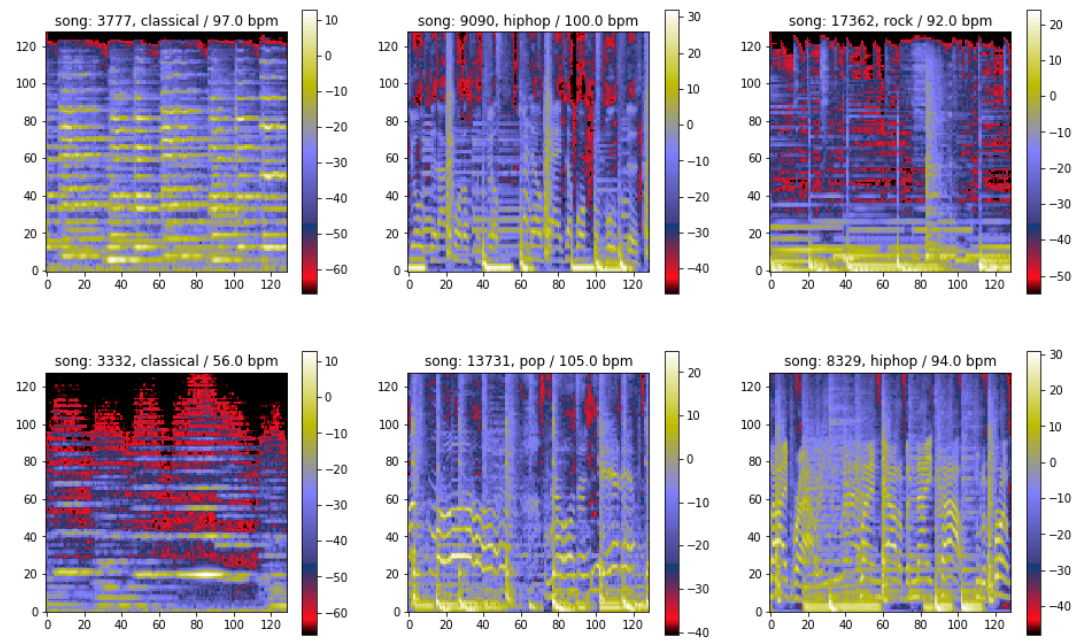


# Metodología I: Datos

## Espectrograma Coeficientes Cepstral (MFCC)



## MFCC con conversión a decibelios



# Metodología II: Meta-parámetros

- Estructura
  - Simple (serie)
  - Paralela (con ramas)
- Filtros
  - Filtros generales (3x3, 5x5, 2x2,...)
  - Filtros horizontales (8x2, 4x1, 14x1,...)
  - Filtros Verticales (2x8,...)
- Otros
  - Optimizador (tipo y ratio inicial de aprendizaje)
  - Activaciones (relu, elu, ...)

# Metodología III: Herramientas



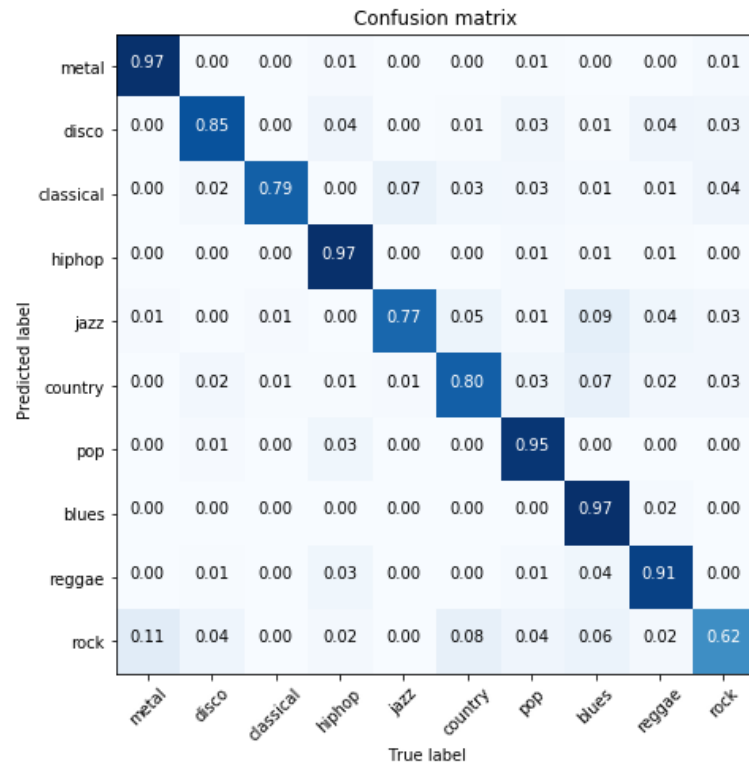
TensorBoard



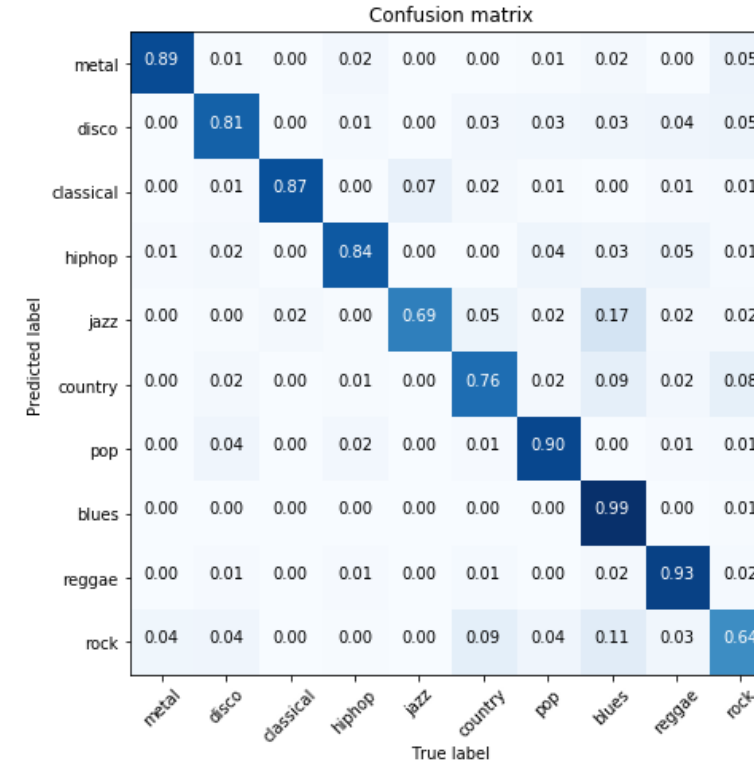
# Resultados

# Resultados I: Predicción de Género

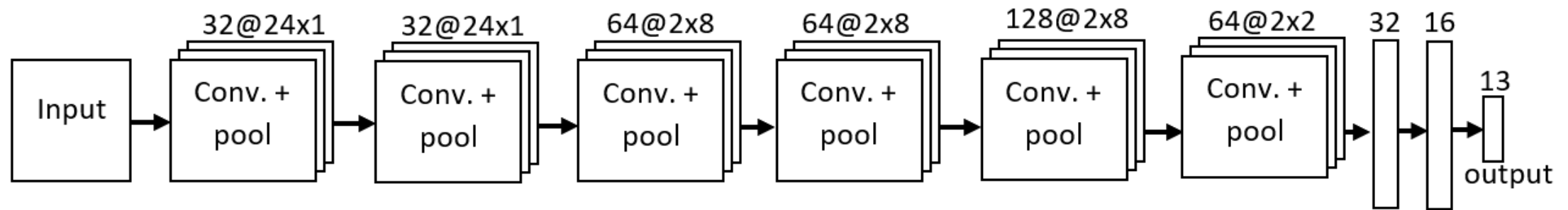
## Arquitectura simple (precisión=0.84)



## Arquitectura paralela (precisión=0.82)



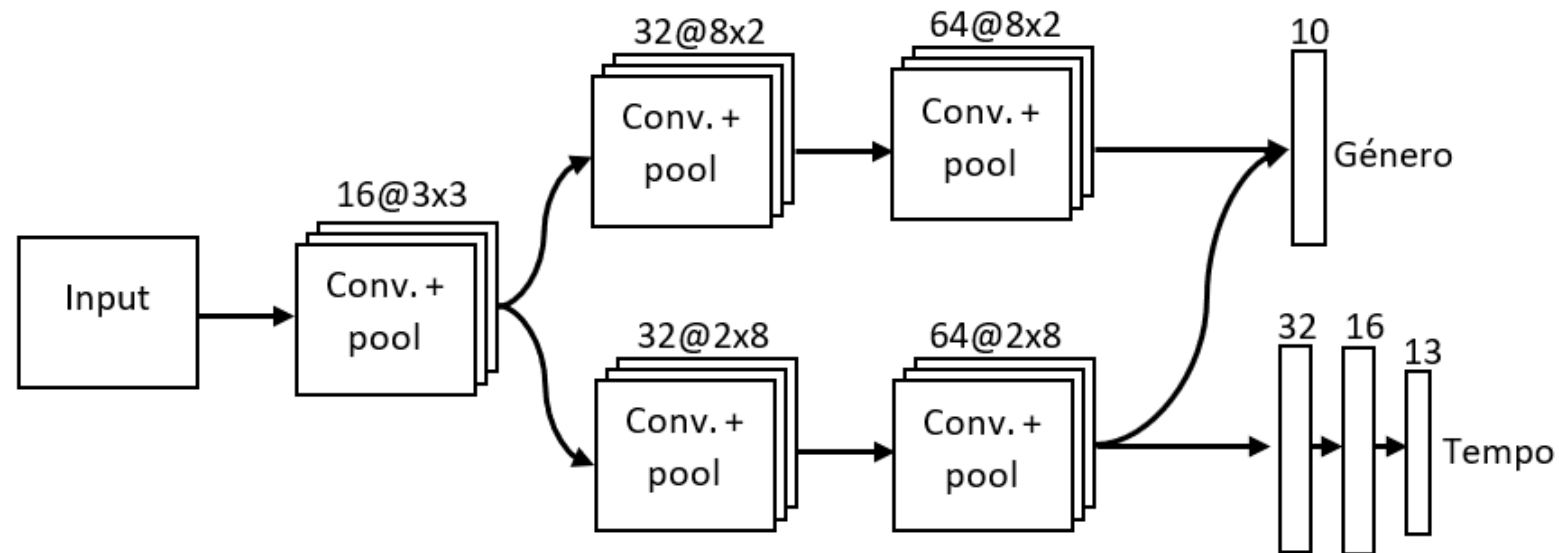
# Resultados II: Predicción de tempo



	Entrenamiento	Test
<b>Red H2V</b>	0.57	0.57
<b>Red V2H</b>	0.56	0.50



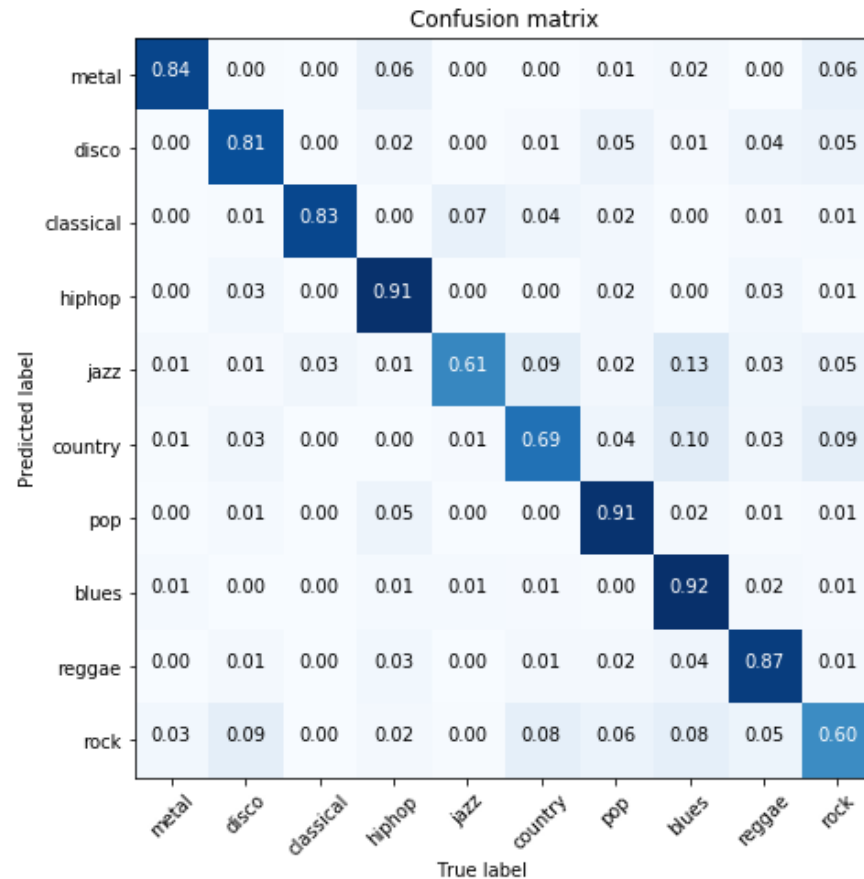
# Resultados III: Predicción simultánea



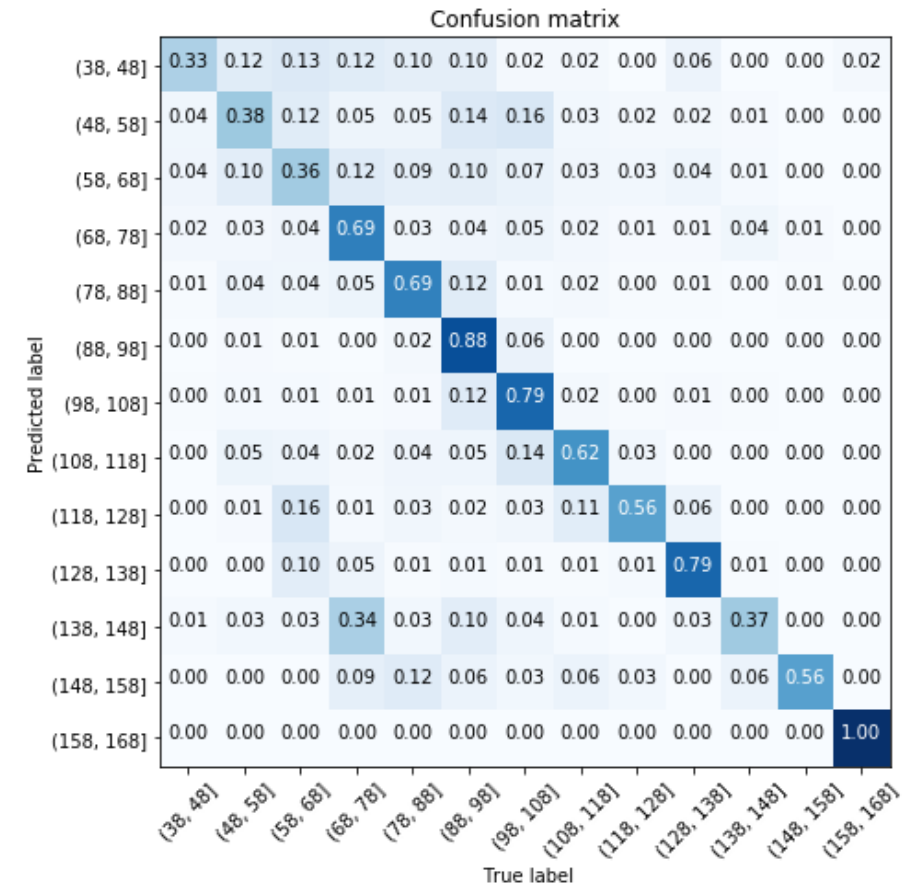
	Entrenamiento	Test
<b>Género</b>	0.82	0.78
<b>Tempo</b>	0.68	0.62

# Resultados III: Predicción simultánea

## Género



## Tempo



# Conclusiones

# Conclusiones

- Extracción de información musical usando CNN
- El género musical se predice con facilidad debido a la textura
- Métricas de tempo son más complicadas pero también posibles
- La estabilidad de predicción género ayuda al entrenamiento con tempo
- Una red compleja sería capaz de predecir ritmos específicos (futuro...)

# Preguntas