

# Curso de aprendizaje profundo

*PCIC, UNAM*

## Tarea 2: redes convolucionales y transferencia

Resuelve los siguientes ejercicios en libretas de Colab que sean replicables<sup>1</sup>. Sugerencias: probar tasas de aprendizaje bajas  $[10^{-3}, 10^{-4}]$ , usar una arquitectura eficiente, cuidar el congelado de parámetros/estadísticas, usar GPU, encapsular y parametrizar funciones de entrenamiento.

### 1. Operación de convolución (2 pts)

Extiende la operación de convolución para imágenes en escala de grises (un solo canal) vista en clase ([https://github.com/gibranfp/CursoAprendizajeProfundo/blob/2024-1/notebooks/2a\\_convolucion.ipynb](https://github.com/gibranfp/CursoAprendizajeProfundo/blob/2024-1/notebooks/2a_convolucion.ipynb)) a imágenes a color (múltiples canales).

### 2. Clasificación de rostros por grupo etario (4 pts.)

Entrena y evalúa modelos de clasificación de rostros por grupo etario usando el conjunto de datos FairFace<sup>2</sup>. Para ello, deberás:

- Dividir aleatoriamente el conjunto de entrenamiento en subconjuntos de entrenamiento y validación y usar el conjunto de validación como prueba.
- Programar tu propia clase para la carga de datos.
- Agregar acrecentamiento de datos.
- Entrenar un modelo similar al de la libreta [https://github.com/gibranfp/CursoAprendizajeProfundo/blob/2024-1/notebooks/2b\\_convnet\\_resnet.ipynb](https://github.com/gibranfp/CursoAprendizajeProfundo/blob/2024-1/notebooks/2b_convnet_resnet.ipynb), pero cambiando los bloques ResNet por bloques ConvNeXt.

---

<sup>1</sup>Todos los ejercicios tienen el mismo peso

<sup>2</sup>Disponible en <https://github.com/joojs/fairface>

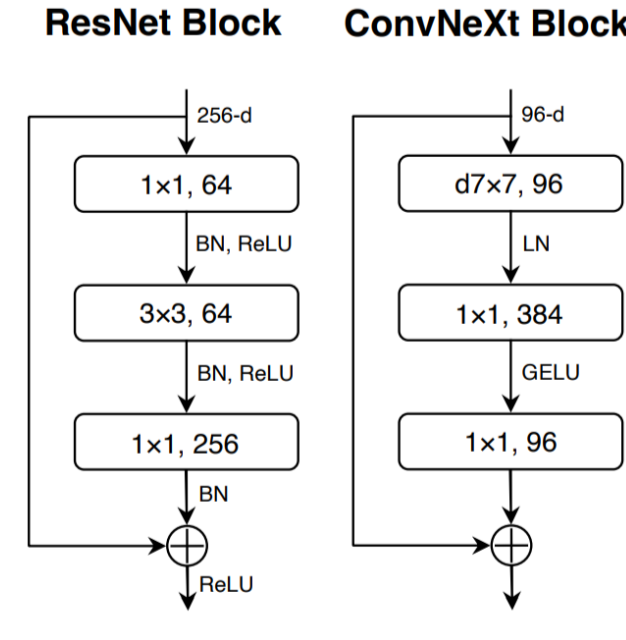


Imagen tomada de Liu et al. A ConvNet for the 2020s, 2022.

- Entrenar un modelo mediante aprendizaje por transferencia a partir de alguna red pre-entrenada actualizando solo los parámetros de la capa de salida y otro modelo actualizando los parámetros de toda la red.
- Graficar las pérdidas y métricas de entrenamiento y validación y discutir todos los resultados.

### 3. Reconocimiento de comandos de voz (4 pts.)

Compara modelos basados en redes recurrentes y convolucionales para el reconocimiento de comandos de voz usando el conjunto Speech Commands<sup>3</sup>. En particular, deberás:

- Entrenar y evaluar un modelo basado en capas convolucionales 1D.
- Entrenar y evaluar un modelo basado en capas convolucionales 2D.
- Graficar las pérdidas y métricas de entrenamiento y validación y discutir todos los resultados.

En todos los casos, se deberán representar los comandos de voz mediante espectrogramas Mel, Mel-Log o MFCCs.

<sup>3</sup>Puedes usar el código para la carga de datos de la libreta [https://github.com/gibranfp/CursoAprendizajeProfundo/blob/2024-1/notebooks/1e\\_procesamiento\\_audio.ipynb](https://github.com/gibranfp/CursoAprendizajeProfundo/blob/2024-1/notebooks/1e_procesamiento_audio.ipynb)