

Avance Semanal del Proyecto ClasifEye

Semana del 15 al 21 de junio de 2025

Gabriela Guzmán Juan Esteban Cardenas Juan David Ardila
Manuel Eduardo Díaz Gabriela Gallegos

21 de junio de 2025

- Semana anterior: Problemas con el preprocesamiento de imágenes.
 - Límite de procesamiento en ~ 600 imágenes debido a restricciones computacionales.
 - Necesidad de extraer etiquetas específicas para entrenar el modelo.
- Objetivo: Optimizar el preprocesamiento y comenzar el entrenamiento de la CNN.
- Estado del proyecto: Dataset inicial preparado, pero con cuellos de botella en el pipeline.

- **Capacidad de procesamiento limitada:**

- El pipeline original no podía manejar más de 600 imágenes sin colapsar.
- Errores esporádicos al cargar imágenes o parsear nombres de archivo.

- **Extracción de etiquetas:**

- Formato inconsistente en nombres de archivo, generando valores "desconocido.^{en} sexo, edad y raza.
- Necesidad de incluir coordenadas faciales (MTCNN keypoints) en el dataset.

- **Preparación para CNN:**

- Pendiente la codificación de etiquetas categóricas (sexo, raza) y normalización de coordenadas.

- **División del dataset en lotes:**

- Dataset dividido en 4 subconjuntos (Gabriela, David, Manuel, Huertas).
- Cada subconjunto con 2 lotes de 600 imágenes para procesamiento paralelo.

- **Optimización del pipeline:**

- Preprocesamiento distribuido en Google Colab para cada lote.
- Generación de CSV con etiquetas: filename, output_{path}, sexo, edad, raza, mtcnn_{confidence}, y coordenadas faciales.

- **Inicio de la CNN:**

- Modelo base con ResNet50 y capas densas para regresión de edad.
- Entrada combinada: características de imágenes (2048) + coordenadas (10).

- **Preprocesamiento completado:**

- 4800 imágenes procesadas ($4 \text{ personas} \times 2 \text{ lotes} \times 600 \text{ imágenes}$).
- Pipeline: reducción de ruido, corrección de iluminación, recorte/alineación facial, normalización.
- CSV generado con etiquetas completas para entrenamiento.

- **Entrenamiento iniciado:**

- Modelo CNN basado en ResNet50 configurado.
- Entrenamiento en curso con $\text{batch_size} = 32$, $\text{epochs} = 10$, $\text{y validaci\,on del } 20\%$.

- **Organización del equipo:**

- Tareas distribuidas para procesar lotes en paralelo.
- Coordinación para estandarizar nombres de archivo y evitar errores.

- **Líder de la próxima semana:** Juan Esteban Cardenas Huertas