

Alumno: Enrique Ramos Diaz

Reporte Problemas encontrados en el dataset

Problemas en Imágenes:

- En las clases de los platillos **verduras salteadas** y **MojarraMojoAjo** existen 59 imágenes.
- En la clase del platillo **pescadoVeracruzana** existen 53 imágenes.
- En la clases de los platillos **arrozRojo**, **chicharronSalsaVerde**, **pechugaasada** y **tlayudas** existen 51 imágenes.
- En la clase del platillo **ChileNogada** sólo hay 49 imágenes.
- En la clase del platillo **esquites** solo hay 46 imágenes.
- Existen 2 carpetas repetidas con imágenes iguales de las clases de los platillos **Pambazos**, **Tostadas**, **pechugaasada**, **verduras salteadas**, **Huevos rancheros** y **Torta ahogada**. Se eliminó una de ellas para cada platillo.
- En las clases de los platillos que tenían más de 50 imágenes, se eliminaron las últimas excedentes.
- En las clases de los platillos que tenían menos de 50 imágenes, se agregaron más imágenes del platillo hasta completarlo.

Problemas en JSONS:

- Los jsons de las clases de los platillos **BurritoArrachera** y **CamaronesEmpanizadosAlCoco** no están en UTF-8. Se guardaron manualmente con esta codificación.
- En el json de la clase del platillo **ChicharronPreparado** hay una coma después de la última llave.
- En el json de clase del platillo **TampiqueñaRes**, el nombre del ingrediente está mal escrito.
- En los json de las clases de los platillos **CamaronesEmpanizadosAlCoco**, **Guacamole**, **MojarraMojoAjo**, **TampiqueñaRes** y **ChongoZamorano** a las llaves le faltan comas.
- En todos los jsons la llave de procedencia esta ubicada dentro del arreglo de ingredientes, a excepcion de las clases de los platillos **arrozRojo**, **PastelAzteca**, **HuevosALaMexicana**, **TamalesOaxaqueños**, **ChongoZamorano**, **chicharronSalsaVerde**, **Huaraches**, **Mixiote**, **MolletesClasicos**, **PozoleVerdeDePollo**, **Sopes** y **TacosAlPastor**.
- En el json de la clase del platillo **ChongoZamorano**, la etiqueta de cada ingrediente estaba en plural.
- En los jsons de las clases de los platillos **TamalesOaxaqueños** y **Guacamole** tenía todo su formato equivocado.
- En los jsons de las clases de los platillos **ajoComino**, **Chilaquilesverdes**, **chilesRellenos**, **esquites**, **Picadillo** y **TacosDorados**, las etiquetas de calorías son incorrectas, fueron guardadas como "kiloCalorias".
- En los jsons de las clases de los platillos **arrozRojo**, **chicharronSalsaVerde**, **Milanesa**, **pozol**, **Sopes**, **TacosAlPastor** y **Tlacoyos**, las etiquetas de cantidad son incorrectas, indica la unidad de medida.
- En el json de la clase del platillo **Pambazos**, el ingrediente "Aceite" no tiene cantidad.

- En el json de la clase del platillo **Milanesa**, el ingrediente "Cebolla" se repite 2 veces.
- En el json de la clase del platillo **Mixiote**, hay una etiqueta extra de "Ingredientes" y su formato de llaves era incorrecto.
- En el json de la clase del platillo **BurritoArrachera** y **CamaronesEmpanizadosAlCoco**, la etiqueta de "Procedencia" decía "Origen".
- En los jsons de la clase de los platillos **ChongoZamorano**, **ChileSerrano**, **Mixiote**, **PastelAzteca**, **PechugaAsada**, **TamalesOaxaqueños** y **verduras salteadas**, las calorías estaban como cadenas en vez de números.
- En el json de la clase del platillo **CochinitaPibil**, el nombre del platillo es "TingaDePollo".
- En el json de la clase del platillo **pezcardoVeracruzano**, el nombre del platillo es "Tlayudas".
- En el json de la clase del platillo **EnchiladasHuastecas**, el nombre del platillo es "chicharronPreparado".
- En los jsons de la clase de los platillos **enfrijoladas**, **esquites**, **Guacamole**, **pozoleRojo** y **Tlacoyos** no existe la etiqueta de "Procedencia".
- Se eliminaron manualmente los acentos y ñ en todos los jsons y se guardaron con codificación UTF-8.

Aplicación de Data augmentation:

Las variantes que se generan por cada imagen válida ya preprocesada son:

- Volteada a la izquierda (espejo).
- Volteada de cabeza.
- Volteada de cabeza espejo (girada a la izquierda).
- Saturación de color a todas las anteriores (3).
- Saturación de color a la imagen original.

Dando un total de 7 variaciones por imagen, más la imagen preprocesada original (8 en total).

Dataset resultante:

- En total los datos se dividen en 53 clases (platillos).
- Cada clase posee 50 imágenes, con un total de 2650 imágenes.
- Si se aplica data augmentation, cada clase posee 400 imágenes, con un total de 21200 imágenes.
- Se pueden aplicar 3 filtros: RGB, escala de grises y bordes.
- El número de canales depende de los filtros aplicados. Si se deja solo en RGB, son 3. Si se aplican los filtros de escala de grises y detección de bordes, son 5 (uno por cada uno, pues ambos constan de 1 canal).
- Se pueden seleccionar dos proporciones: rectangular o cuadrada.
- Se puede aplicar o no data augmentation.
- Las dimensiones de las imágenes resultantes es 160x112 en proporción rectangular, y 112x112 en proporción cuadrada.
- Se utilizó reshape y rescale para disminuir las dimensiones originales de las imágenes (rectangular: 640x448. cuadrado: 448x448).
- En total se generaron 4 archivos pickles de 2GB cada uno.
- Las clases de los platillos fueron codificadas en onehot, como arreglos de numpy float32.

- Las imágenes fueron convertidas a arreglos de numpy float32.
- Estos arreglos de las imágenes fueron normalizados (divididos entre 255).
- Los jsons fueron guardados como diccionarios.
- Al final resultan dos listas: una con clase onehot + imagen, y otra con clase onehot + json.