# Super Resolucion con Autoencoders Convolucionales

Armando Rios-Lastiri

No Institute Given

**Resumen** La resolución es una característica importante para determinar la naturaleza y las características de la imagen. Mejorar la resolución fortalece las características ocultas dentro de la imagen y hace que la imagen sea más nítida e informativa. Se desarrollo una arquitectura de red neuronal convolucional (CNN) agregando diferentes capas a la red neuronal. Se propuso un autocodificador capaz de codificar y decodificar la estructura de las imágenes para mejorar su resolución. El modelo aprende las características de menor dimensión de las imágenes poco claras y les proporciona una alta resolución al predecir y mejorar sus dimensiones. El modelo se entreno en imágenes de baja resolución y las imágenes de alta resolución correspondientes, y se implemento un codificador convolucional para eliminar el ruido de la imagen e introducir alta resolución en las imágenes borrosas o corruptas.

## 1. Introducción

El objetivo central de la superresolución (SR) es generar una imagen de mayor resolución a partir de imágenes de menor resolución. La imagen de alta resolución ofrece una alta densidad de píxeles y, por lo tanto, más detalles sobre la escena original. La necesidad de alta resolución es común en las aplicaciones de visión por computadora para un mejor desempeño en el reconocimiento de patrones y análisis de imágenes. La alta resolución es de importancia en las imágenes médicas para el diagnóstico. Muchas aplicaciones requieren hacer zoom en un área específica de interés en la imagen en la que la alta resolución se vuelve esencial, p. Ej. aplicaciones de vigilancia, forense y de imágenes por satélite.

## 2. Contexto del problema

La superresolución se basa en la idea de que se puede utilizar una combinación de secuencias de imágenes de una escena de baja resolución (ruidosas) para generar una imagen de alta resolución o una secuencia de imágenes. Por lo tanto, intenta reconstruir la imagen de la escena original con alta resolución dado un conjunto de imágenes observadas a menor resolución. El enfoque general considera las imágenes de baja resolución como resultado del remuestreo de una imagen de alta resolución. Entonces, el objetivo es recuperar la imagen de alta resolución que, cuando se muestrea nuevamente en función de las imágenes de entrada y el modelo de imagen, producirá las imágenes observadas de baja resolución.

## 3. Trabajo relacionado

Para este trabajo se hizo uso de la base de datos llamada

## 4. Marco teorico

Single image super resolution (SISR) se refiere a la reconstrucción de una imagen de alta resolucion (HR) a partir de una observación de baja resolucion (LR). Dada una imagen LR $\mathbf{Y}$, generalmente se supone que está degradada a partir de una imagen HR correspondiente X, que se puede representar como

$$\mathbf{Y} = D\left(\mathbf{X}, \theta_D\right)$$

donde $D(\cdot)$ denota el proceso de degradación definido por el conjunto de parámetros $\theta_D$. En un escenario real, el parámetro de degradación $\theta_D$ es desconocido, y todo lo que tenemos es la imagen LR $\mathbf{Y}$. SISR tiene como objetivo recuperar una buena estimación de la imagen de HR potencial mediante la inversión del proceso de degradación que se muestra en la Ec. (1), que puede formularse como

$$\hat{\mathbf{X}} = R\left(\mathbf{Y}, \theta_R\right)$$

donde $R(\cdot)$ representa la función SR y $\theta_R$ es el conjunto de parámetros correspondiente. $\hat{X}$ es la imagen superesuelta de $\mathbf{Y}$, es decir, una estimación de la imagen HR real $\mathbf{X}$.

Aparentemente, el proceso de SR y el proceso de degradación son inversos entre sí. Por lo tanto, para obtener un excelente rendimiento de reconstrucción, la función SR $R\left(\mathbf{Y}, \theta_R\right)$ debe adaptarse a la degradación $D\left(\mathrm{X}, \theta_D\right)$. En la literatura, algunos investigadores [36] − [45] aproximan la degradación a través del desenfoque, reducción de resolución e inyección de ruido. Matemáticamente, el proceso de degradación simulado es el siguiente

$$\mathbf{Y} = \mathbf{SBX} + \mathbf{n}$$

donde $\mathbf{B}$ y $\mathbf{S}$ denotan las operaciones de desenfoque y reducción de resolución, respectivamente. En general, el desenfoque se realiza convolucionando la imagen HR con un kernel gaussiano. n representa el ruido aditivo, que generalmente se supone que es ruido blanco gaussiano o se puede ocupar un modelo de degradación más simple, es decir, reduciendo directamente la escala de una imagen HR utilizando el núcleo "bicúbico"para generar la imagen LR correspondiente, este ultimo metodo fue el que se implemento en este proyecto.

## 5. Descripción del proyecto

### 5.1. Exploracion de los datos

CelebFaces Attributes Dataset (CelebA) es un conjunto de datos de atributos faciales a gran escala con más de 200.000 imágenes de celebridades. Las imágenes

de este conjunto de datos cubren grandes variaciones de pose y desorden de fondo. El conjunto de datos cuenta con 202,599 de imágenes con una resolucion nativa de 178x128 pixeles.

Para la particion de los datos se usaron 1000 imagenes para prueba, 40,320 para evaluacion y 161,279 para el entrenamiento.

### 5.2. Preparación de los datos

Para el preprocesamiento de las imágenes primero se normalizaron cada una de ellas despues, se recortaron a 120x120 píxeles cada una, este conjunto son las imágenes HR (X). Para obtener las imágenes LR (Y) se escalaron las imágenes a una resolución de 45x45 píxeles, posteriormente se volvieron a escalar a 120x120 píxeles para así obtener las imágenes con ruido.

### 5.3. Métodos y modelos

Para el codificador se ocuparon dos bloque convolucionales, cada uno consta de dos capas convolucionales y un Max pool com filtros de 3x3 y funcion de activacion ReLu, estos bloques se pasan a una ultima capa convoluacional para asi formar el codificador completo.

El decodificador consta de dos bloques cada uno con una cada de sobre muestreo y dos capas convoluciones, estos bloques pasan a una ultima capa convolucional para asi conformar el decodificador.

Por ultimo al modelo se le agregaron dos conexiones residuales uno entre la capa 5 y la 9 y otro entre la capa 2 y 13. El modelo cuenta con 1,110,403 parametros.

### 5.4. Evaluación de los modelos

Para aprender el mapeo de imágenes corruptas a imágenes limpias se necesita estimar los pesos $\Theta$ representados por los núcleos convolucional y deconvolucional. Específicamente, dada una colección de $N$ pares de muestras de entrenamiento $\{X^i, Y^i\ right\}$, donde $X^i$ es una imagen ruidosa y $Y^i$ es la versión limpia como la verdad básica. Minimizamos el siguiente error cuadrático medio (MSE):

$$\mathcal{L}(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \left\| \mathcal{F}\ left(X^i; \Theta\right) - Y_F^{i\,2} \right.$$

El modelo se entreno durante 10 epocas y se ocupo un optimizador Adam y la metrica de accuracy. Los resultados del entrenamiento se muestran en la siguiente figura.

## 6. Conclusiones y recomendaciones

El uso de autoencoders ha mostrado ser un buen metodo para mejorar la resolcuciones de las imagenes de rostros con las que fue entrenado. La funcion de perdida MSE

## Referencias