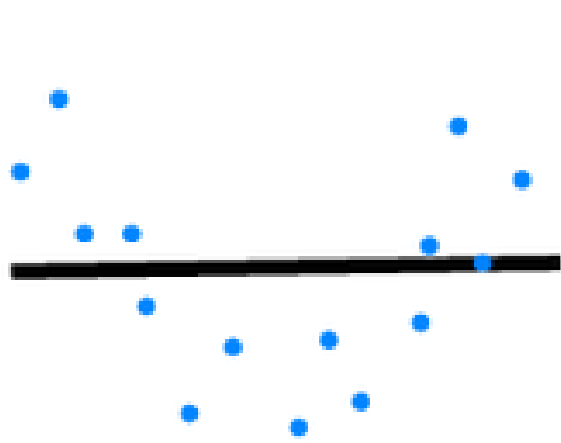
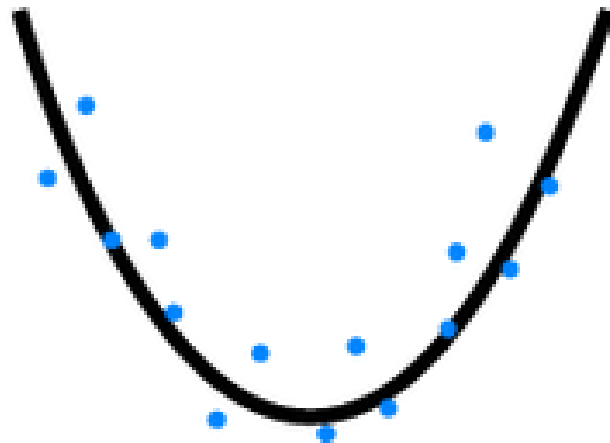


Regularización

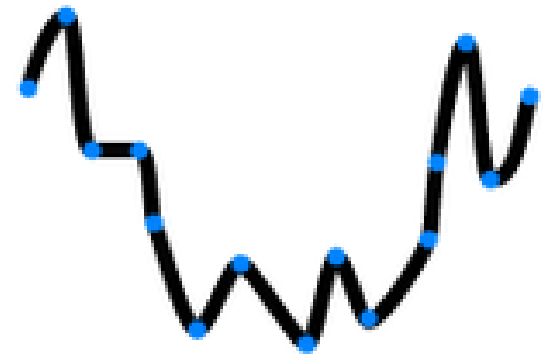
Overfitting



Underfitting

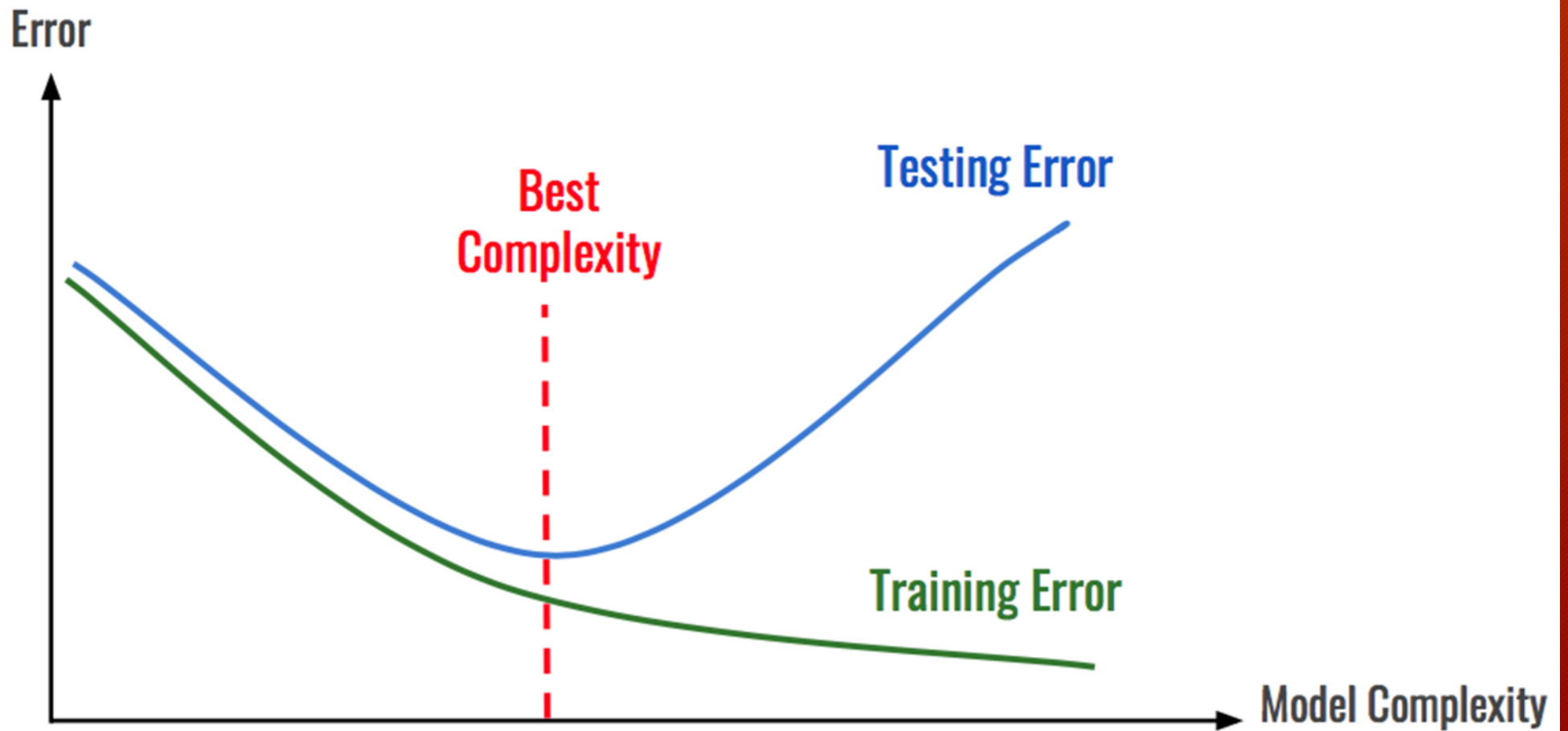


Desired



Overfitting

Overfitting



Regularización

- Existen muchos w que pueden minimizar una función de pérdida para un dataset de entrenamiento dado.
- ¿Hay alguno de ellos que beneficie más que el resto la generalización de nuestro modelo?
- Pensemos en un perceptrón simple: una idea es buscar aquel w que evite las predicciones estén demasiado influenciadas por sólo algunas dimensiones de entrada, repartiendo esa responsabilidad.
- Esto implica evitar que existan w_i muy grandes.

L2 Weight Decay

Agregar un término de regularización a la función de pérdida que sea la norma euclídea cuadrada de los pesos

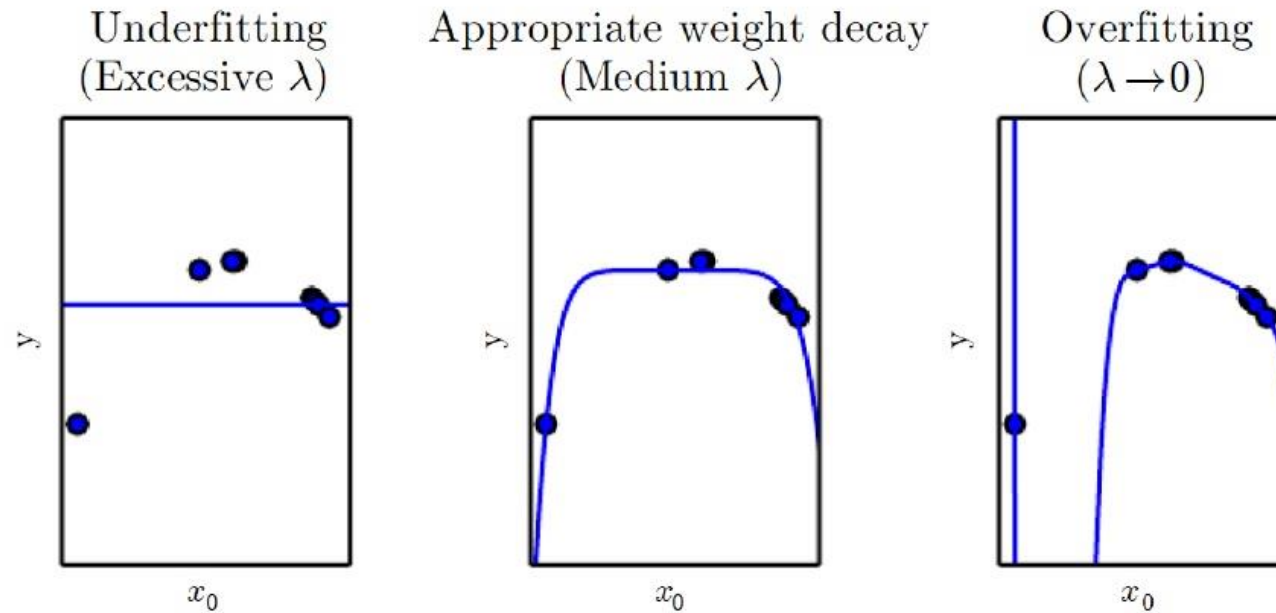
$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \frac{1}{N} \sum_{n=1}^N \mathcal{L}((\mathbf{x}, d)_n; \mathbf{w}) + \frac{1}{2M} \lambda \underbrace{\|\mathbf{w}\|_2^2}_{\substack{\|w\|_2^2 = \left(\left(\sum_{m=1}^M w_m^2 \right)^{1/2} \right)^2 = \sum_{m=1}^M w_m^2}}$$

N = Número de data samples en el training set

M = Número de parámetros en w

L2 Weight Decay

Influencia del parámetro λ : Ejemplo al fitear un polinomio de orden 9 sobre datos sampleados de una función cuadrática

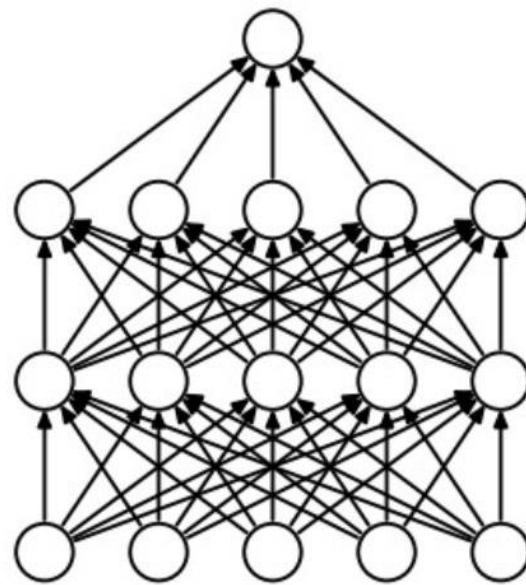


Drop - Out

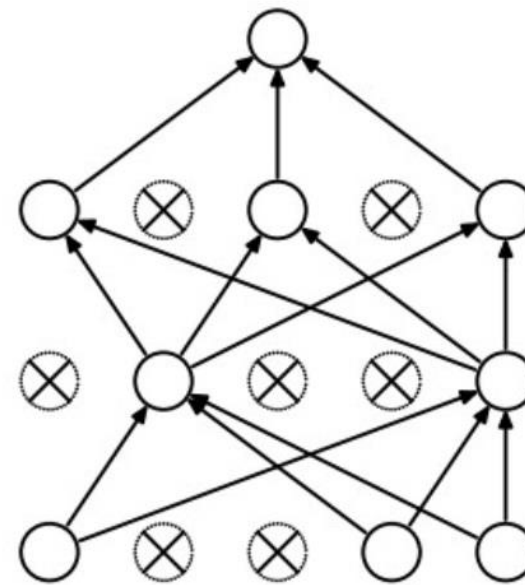
- **Intuición:** combinar las predicciones de múltiples modelos entrenados con el mismo fin es una forma de prevenir overfitting.
- **Idea:** promediar las predicciones de múltiples modelos entrenados independientemente para resolver el mismo problema.
- **Problema:** ¡Es carísimo!
- **Solución:** hacer que un mismo modelo sea a la vez "múltiples" modelos.

Drop - Out

Aleatoriamente, con un probabilidad $(1-p)$, las neuronas son ignoradas en la pasada forward durante entrenamiento



(a) Standard Neural Net



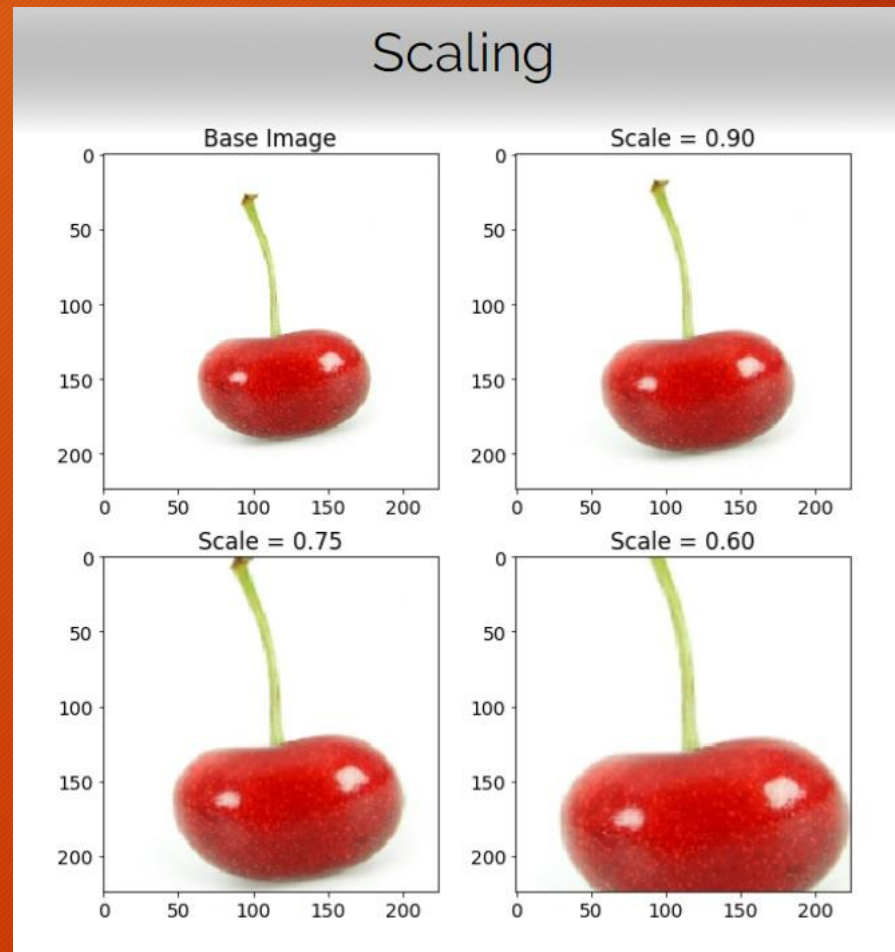
(b) After applying dropout.

En test time, todas las neuronas son utilizadas y los pesos de salida de la neurona son multiplicados por p

Aumentación de Datos

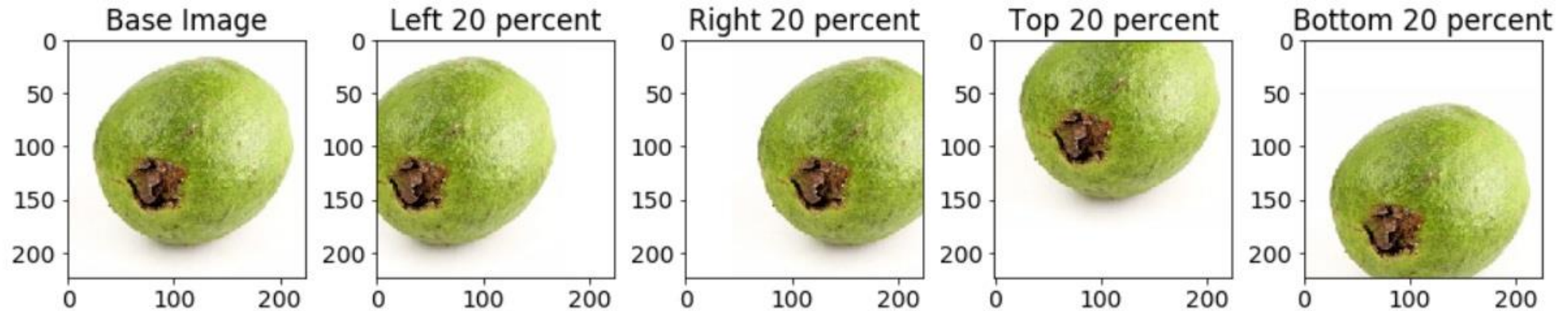
- En tiempo de entrenamiento, aumentar artificialmente el dataset utilizando transformaciones en los datos y conservando las etiquetas.
- En tiempo de prueba, es posible generar N versiones de la imagen de test.

Aumentación de Datos



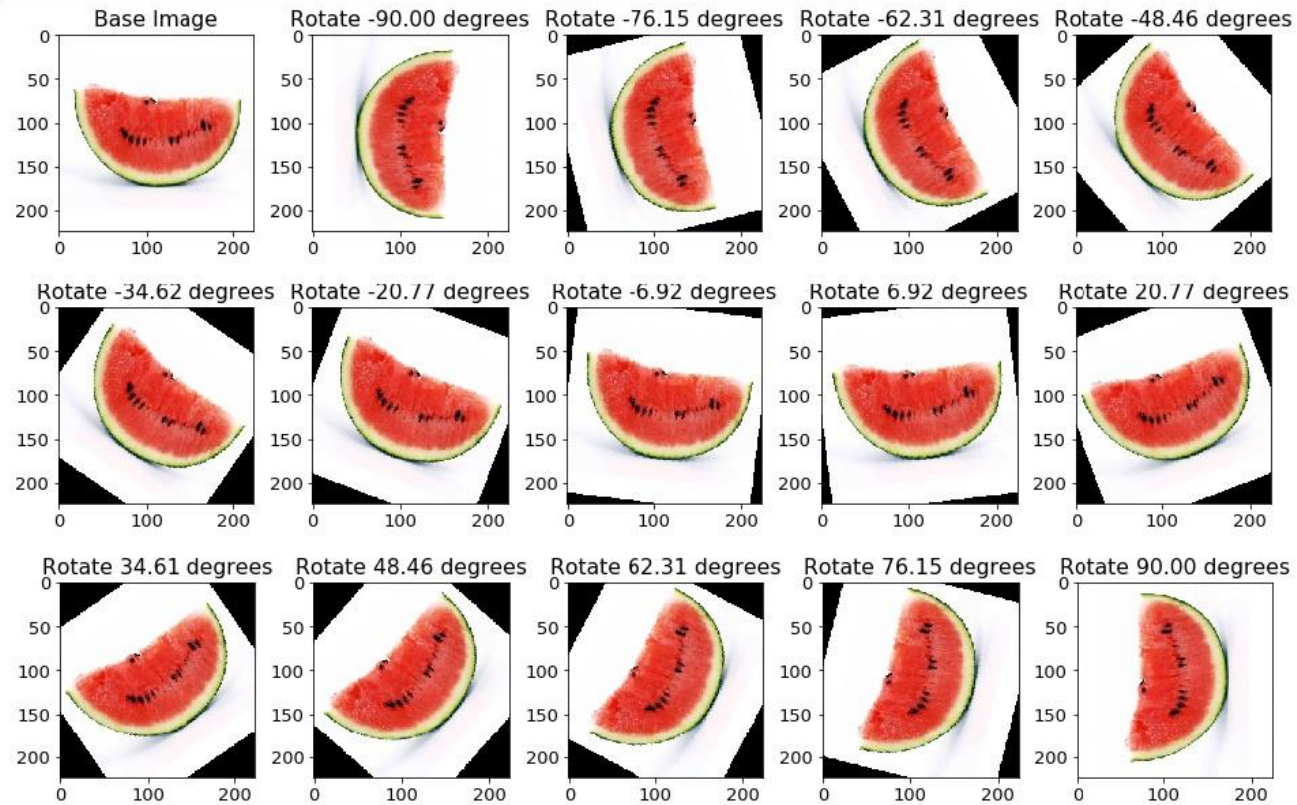
Aumentación de Datos

Translation

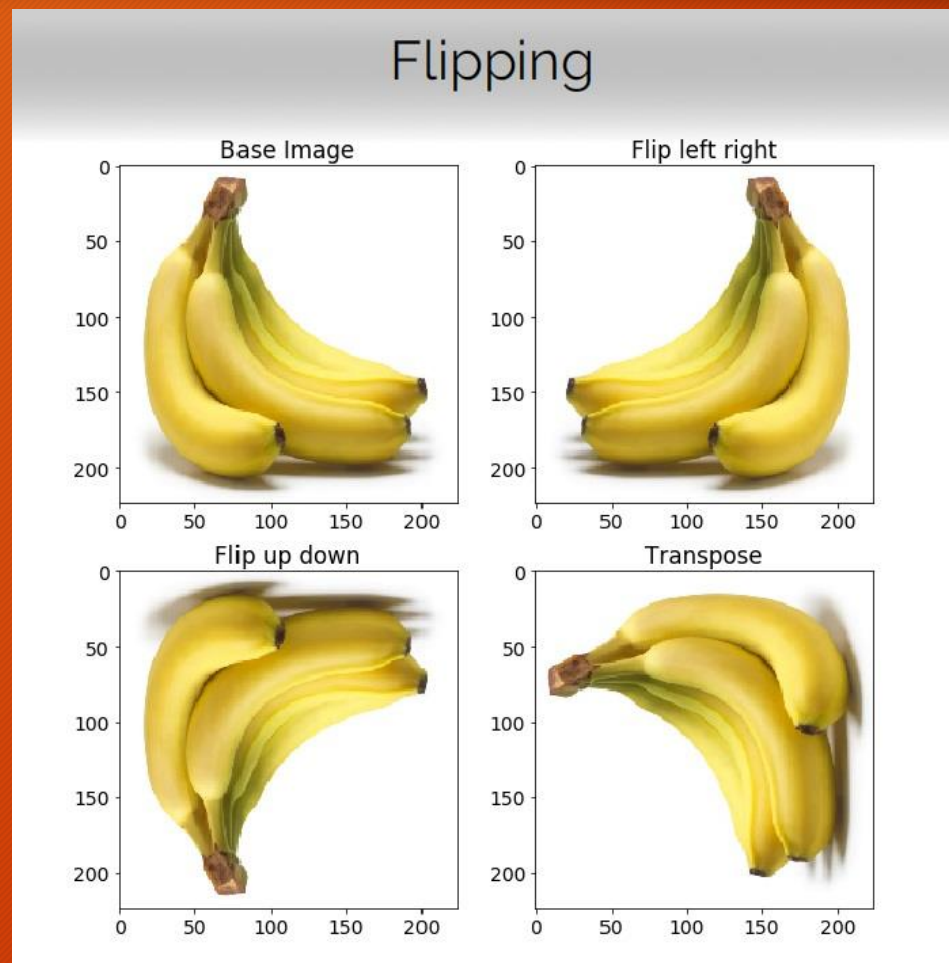


Aumentación de Datos

Rotation

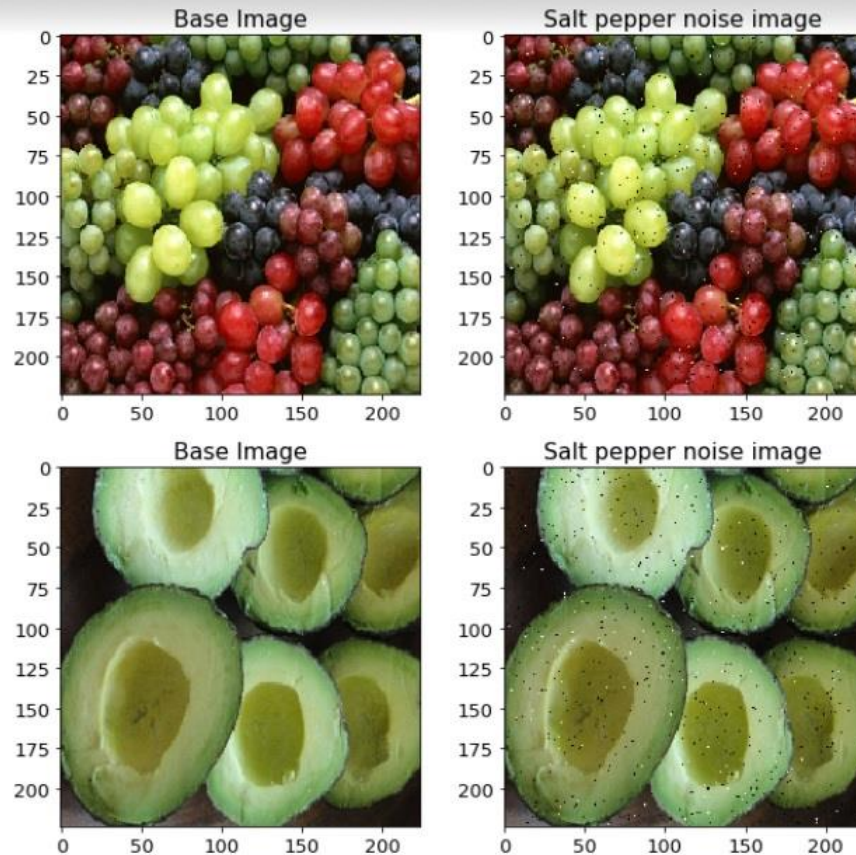


Aumentación de Datos



Aumentación de Datos

Ruido



Bibliografía

- Taller: “Aprendizaje Profundo para el análisis de imágenes biomédicas”. Profesor Dr. Enzo Ferrante. Bloque 2. UNL, Santa Fe, Argentina.