

# Transfer Learning in Image Classification

## Taller III

June 4, 2024

## 1 Inception V3 Model

### 1.1 Arquitectura

El modelo Inception V3 es una versión avanzada y optimizada del modelo Inception V1, diseñada para mejorar la adaptación del modelo mediante varias técnicas de optimización. Este modelo es más eficiente, tiene una red más profunda en comparación con Inception V1 y V2 sin comprometer la velocidad, es computacionalmente menos costoso y utiliza clasificadores auxiliares como regularizadores. Lanzado en 2015, el Inception V3 tiene un total de 42 capas, aproximadamente 23.9 millones de parámetros y una tasa de error menor que sus predecesores. Las principales optimizaciones incluyen la factorización de convoluciones grandes en más pequeñas, la factorización espacial en convoluciones asimétricas, el uso de clasificadores auxiliares para mejorar la convergencia en redes muy profundas, y la reducción eficiente del tamaño de la cuadrícula. Estas técnicas resultaron en un modelo con mejor precisión y menores costos computacionales.

### 1.2 Resultados

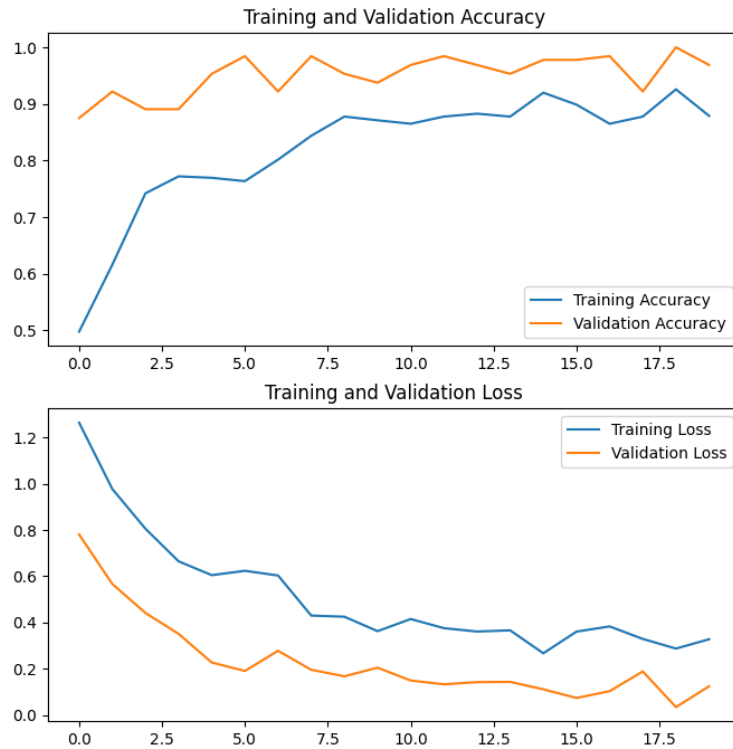


Figure 1: Training and Validation Accuracy and Loss in InceptionV3 Model

## 2 MobileNet Model

### 2.1 Arquitectura

La arquitectura de MobileNet incluye 27 capas de convolución, que comprenden 13 convoluciones profundas (depthwise), 1 capa de promedio (Average Pool), 1 capa completamente conectada y 1 capa Softmax. Específicamente, tiene 13 convoluciones profundas de 3x3, 1 convolución de 3x3 y 13 convoluciones de 1x1, siendo estas últimas las que consumen el 95% del tiempo de procesamiento. Este modelo, desarrollado por Andrew G. Howard y otros investigadores de Google, cuenta con 4.2 millones de parámetros en su versión estándar, mientras que las versiones más pequeñas tienen 1.32 millones de parámetros, lo cual es bajo en comparación con otros modelos de aprendizaje automático estándar.

### 2.2 Resultados

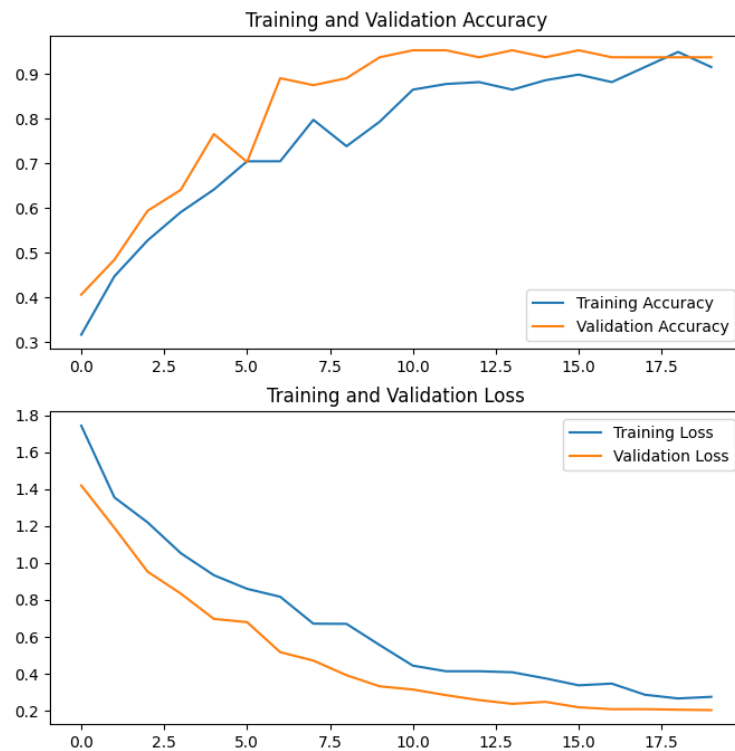


Figure 2: Training and Validation Accuracy and Loss in MobileNet Model

### 3 InceptionResNetV2 Model

#### 3.1 Arquitectura

El modelo InceptionResNetV2 es una combinación de las arquitecturas Inception y ResNet, diseñada para mejorar la eficiencia y precisión del modelo. Lanzado en 2016, este modelo combina los bloques Inception con conexiones residuales, lo que ayuda a mitigar el problema de la degradación en redes muy profundas y facilita la optimización del entrenamiento. InceptionResNetV2 tiene un total de 164 capas y aproximadamente 55.9 millones de parámetros, lo que le permite capturar patrones complejos en los datos con alta precisión. Las conexiones residuales permiten entrenar redes más profundas sin perder información crucial, mientras que los módulos Inception mejoran la capacidad del modelo para procesar diferentes escalas de características. Esta combinación de técnicas resulta en un modelo altamente eficiente, con una mayor capacidad para generalizar y un menor costo computacional en comparación con modelos anteriores.

#### 3.2 Resultados

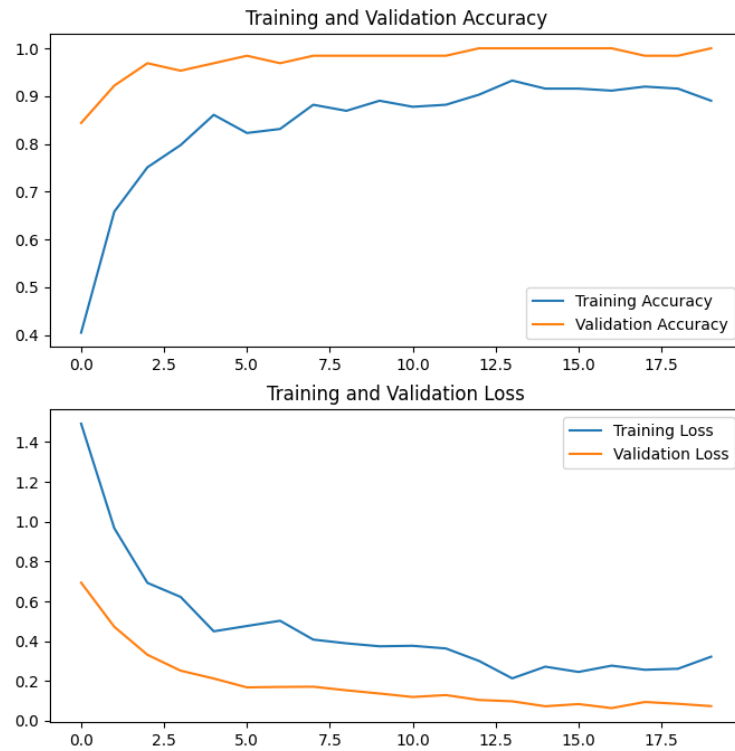


Figure 3: Training and Validation Accuracy and Loss in InceptionResNetV2 Model

## 4 Comparativa y Conclusiones

### 4.1 Matrices de Confusión

Ahora bien, para cada uno de los tres modelos contruiremos las matrices de confusión en los dos escenarios:  
i) usando el mejor resultado predicho y ii) usando los dos mejores resultados predichos:

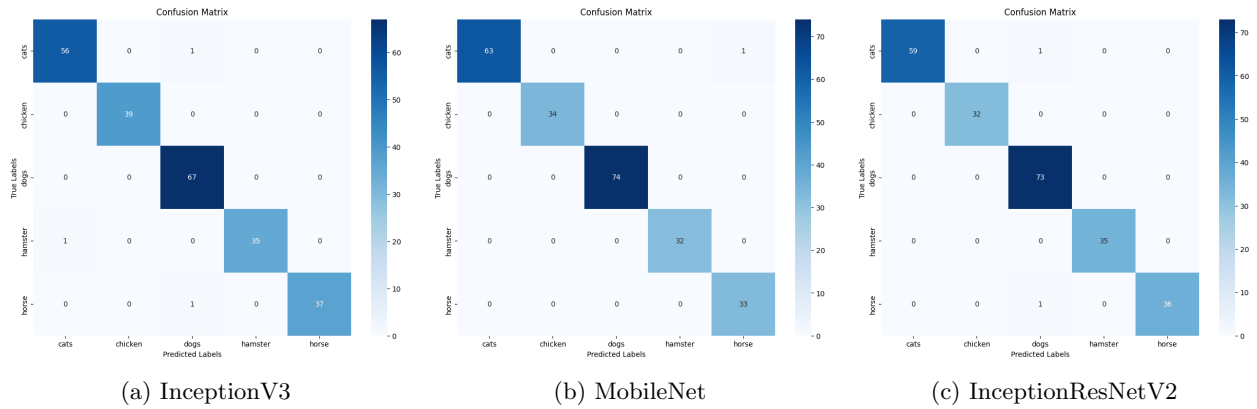


Figure 4: Comparison of Confusion Matrices top one result

Al analizar el performance con cada una de las distintas redes para la predicción convencional, notamos que los resultados generales para las tres redes es muy alto, algo sorprendente para la diferencia de layers y parámetros de las tres. No obstante, el mejor accuracy fue el de la red menos densa: MobileNet con 99.15%. En segundo lugar quedó la InceptionResNetV2 con 99.58%, lo cual sorprende al ser la red con el mayor número de parámetros. En tercer lugar, pero muy cercano a las otras fue el modelo InceptionV3 con un accuracy del 98.73%

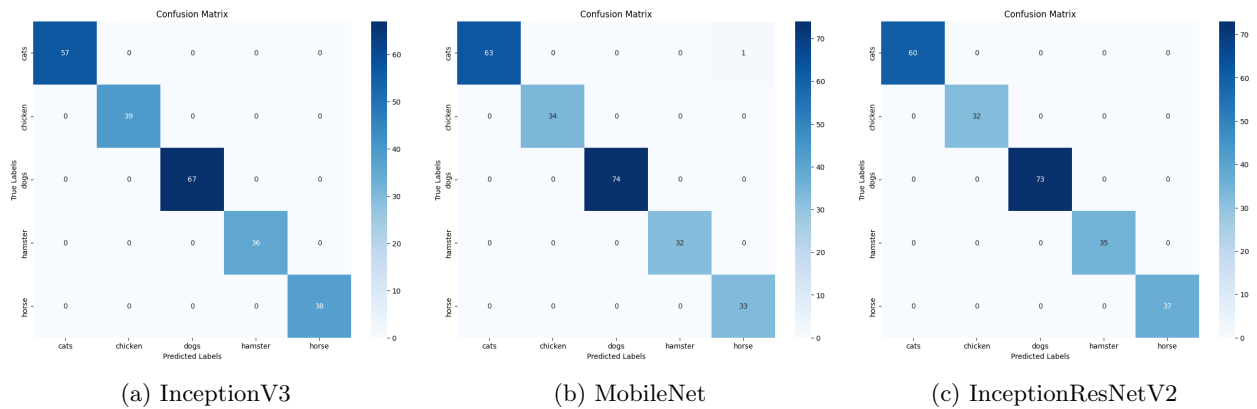


Figure 5: Comparison of Confusion Matrices top two result

A diferencia del caso anterior, los resultados con el dataset de validación usando los dos mejores resultados para los modelos de InceptionV3 e InceptionResNetV2 fueron perfectos en todas las métricas. En cambio, para el modelo de MobileNet solamente tuvo un error de clasificación, que sigue siendo una mejoría al caso anterior y un resultado más alto en todas las métricas al hacer la predicción con solo el mejor resultado.

## 4.2 Comentarios Finales

En este estudio, hemos analizado y comparado tres modelos de redes neuronales convolucionales populares: Inception V3, MobileNet e InceptionResNetV2, en el contexto de la clasificación de imágenes. Cada uno de estos modelos presenta sus propias características arquitectónicas y parámetros, lo que los hace adecuados para diferentes aplicaciones y requisitos computacionales. Sin embargo, al observar los resultados de las matrices de confusión en el conjunto de validación, donde vemos una discrepancia entre el rendimiento del entrenamiento y la validación, es posible que exista un problema de sobreajuste (overfitting) en el proceso de entrenamiento. Esto sugiere la necesidad de aplicar técnicas de regularización o ajustar los hiperparámetros del modelo para mejorar su capacidad de generalización y evitar el sobreajuste en futuros experimentos.

## 5 Bibliografía

- OpenGenus Foundation. (n.d.). MobileNet Architecture. *OpenGenus IQ*. Recuperado el 3 de junio de 2024, de <https://iq.opengenus.org/mobilenet-architecture/>
- OpenGenus Foundation. (n.d.). Inception V3 model architecture. *OpenGenus IQ*. Recuperado el 3 de junio de 2024, de <https://iq.opengenus.org/inception-v3-model-architecture/>
- OpenGenus Foundation. (n.d.). Inception-ResNet V1. *OpenGenus IQ*. Recuperado el 3 de junio de 2024, de <https://iq.opengenus.org/inception-resnet-v1/>