Entrenamiento, optimización y evaluación de redes neuronales para clasificación de imágenes

# Introducción

El núcleo de esta investigación gira en torno a tres ejes principales: la optimización de hiperparámetros, la exploración de distintas arquitecturas de red y el entrenamiento completo o fine-tuning de modelos pre-entrenados. Estos aspectos son fundamentales para entender la eficacia y eficiencia de las redes neuronales en la tarea de clasificación de imágenes.

En la primera fase del estudio, se realiza una exhaustiva experimentación con modelos de red como LeNet y AlexNet, aplicados al conjunto de datos Fashion-MNIST. Este análisis permite comprender cómo distintos hiperparámetros, como la tasa de aprendizaje, el momentum, los optimizadores, el dropout, la regularización y el tamaño del lote, influyen en el proceso de entrenamiento de los modelos. La segunda fase del trabajo se centra en el fine-tuning de un modelo de ResNet con el conjunto de datos CIFAR-10, comparando los resultados con el uso directo del modelo ResNet. Esta comparativa ofrece una visión clara sobre la eficacia del fine-tuning en modelos pre-entrenados.

La metodología adoptada incluye la lectura y preprocesamiento del corpus de imágenes, la partición de los datos para entrenamiento, validación y test, seguida por el entrenamiento y optimización de la red. Finalmente, se evalúa el clasificador sobre el conjunto de test con el modelo ajustado de forma óptima. Este enfoque metodológico se apoya en cuadernos y materiales proporcionados en las sesiones prácticas y teóricas del curso.

# Entrenamiento de redes neuronales para imágenes

El presente apartado se enfoca en el entrenamiento de redes neuronales utilizando el corpus Fashion-MNIST, con el objetivo de explorar y comprender la efectividad de diferentes arquitecturas en la clasificación de imágenes. Los experimentos se centran en dos arquitecturas significativas: LeNet y AlexNet.

Por un lado, LeNet ha sentado las bases para muchos desarrollos posteriores en el campo. Su arquitectura simple pero eficaz la convierte en un punto de partida ideal para comprender los fundamentos del entrenamiento de redes neuronales en tareas de clasificación de imágenes mientras que, AlexNet, es notablemente más profunda y compleja que LeNet. Su éxito en la competencia ImageNet ha demostrado su capacidad para manejar tareas de clasificación de imágenes más desafiantes y variadas, lo que la hace relevante para nuestro estudio.

En este apartado, se detallarán los procesos de entrenamiento de ambas arquitecturas, evaluando su desempeño y eficacia en el corpus Fashion-MNIST. El análisis incluirá la optimización de hiperparámetros, la adaptación de la arquitectura a las especificidades del conjunto de datos y la evaluación de los modelos en términos de precisión y eficiencia. Esta exploración proporcionará insights valiosos sobre las fortalezas y limitaciones de cada arquitectura.

## Corpus Fashion-MNIST

El corpus Fashion-MNIST se ha establecido como un estándar de facto para la evaluación de algoritmos de clasificación de imágenes. Este conjunto de datos, que comprende 70,000 imágenes en escala de grises de 10 categorías de prendas de vestir, representa un desafío más idóneo para algoritmos modernos en comparación con el conjunto de datos MNIST tradicional. Con imágenes de 28x28 píxeles, ofrece una complejidad suficiente para probar la eficacia de diferentes arquitecturas de redes neuronales sin ser prohibitivamente grande para la experimentación en entornos con recursos computacionales limitados.

El proceso de experimentación comenzó con la carga y preparación de los datos. Se accedió al conjunto de datos almacenado previamente y se dividió en tres subconjuntos: entrenamiento, validación y prueba. Se reservó el 80% de las imágenes para el entrenamiento y el 20% restante se dividió equitativamente entre validación y prueba. Esta estrategia de partición asegura que el modelo no solo aprenda eficazmente, sino que también se valide y se pruebe de manera que sea representativa de su capacidad para generalizar a nuevos datos.

Tras la carga, los datos fueron normalizados para asegurar que los valores de los píxeles estén en el rango [0, 1], lo que mejora la estabilidad y el rendimiento del entrenamiento de la red. Este preproceso es un paso crítico que impacta directamente en la convergencia del algoritmo durante el entrenamiento.

Con los datos cargados y preprocesados, se procedió a visualizar muestras del conjunto de entrenamiento para confirmar su integridad y entender la naturaleza de las imágenes con las que se trabajaría. La visualización de estas imágenes revela la variedad de prendas y la claridad con la que deben ser clasificadas, estableciendo así una base para la evaluación del desempeño de los modelos de redes neuronales, en la Ilustración 1 podemos apreciar una muestra de ello.

Texto

Descripción generada automáticamente

Ilustración - Muestra de algunos datos del conjunto Fashion-MNIST

Por último, se realizó un resumen estadístico que proporciona una comprensión cuantitativa del conjunto de datos. El conjunto de entrenamiento consta de 48,000 imágenes, el de validación de 9,600, y el conjunto de pruebas alcanza las 10,000 imágenes. Estas cifras no solo garantizan una cantidad suficiente de datos para el entrenamiento robusto de los modelos sino también para su validación y prueba exhaustivas.

## LeNet

El experimento con la arquitectura LeNet en el corpus Fashion-MNIST fue planeado y ejecutado para evaluar el impacto de los hiperparámetros en el rendimiento de la clasificación de imágenes. El modelo LeNet, un pionero entre las redes neuronales convolucionales se seleccionó por su estructura simple y efectividad demostrada en tareas similares.

Para construir el modelo, se empleó la API secuencial de Keras, que permite apilar capas de forma intuitiva y lineal. La red se compuso de capas convolucionales alternadas con capas de pooling, seguidas de capas de dropout para la regularización y capas densas para la clasificación final. Los hiperparámetros clave como la tasa de aprendizaje, el tamaño del batch y la regularización L1 y L2 se variaron para observar sus efectos.

La función de pérdida escogida fue la entropía cruzada categórica esparsa, un estándar para problemas de clasificación multiclase, y se utilizó el optimizador SGD (Descenso de Gradiente Estocástico) para la actualización de pesos. Se implementó también un callback de parada temprana para evitar el sobreentrenamiento, deteniendo el entrenamiento si el rendimiento de validación no mejoraba después de un número definido de épocas.

La experimentación varió las tasas de aprendizaje entre 0.05 y 0.1, los tamaños de batch entre 8, 64 y 128, y la regularización L1 y L2 tanto individualmente como en combinación, estableciendo en 0.0 y 0.001 sus respectivos valores. Esto se hizo con el fin de identificar la configuración óptima que maximizara la precisión de la red mientras minimizaba la posibilidad de sobreajuste.

Se iteró sobre las combinaciones de hiperparámetros utilizando bucles anidados, entrenando y almacenando cada modelo con su configuración particular. Este enfoque no solo facilita un análisis comparativo entre los diferentes modelos, sino que también asegura una reproducción fácil de los experimentos.

La flexibilidad en la definición de la arquitectura LeNet y la selección de hiperparámetros reflejan una experimentación profunda y considerada. Al ajustar y evaluar diferentes combinaciones, se pudo investigar cómo cada hiperparámetro influía en el aprendizaje y la generalización del modelo en el conjunto de datos Fashion-MNIST, proporcionando así una base sólida para la comprensión y mejora de modelos de clasificación de imágenes con redes neuronales convolucionales.

### Resultados

La experimentación con diferentes configuraciones del modelo LeNet ha generado una variedad de resultados, destacando el efecto significativo que los hiperparámetros tienen en el rendimiento del modelo. Los modelos se evaluaron utilizando la métrica F1-score, que es el promedio armónico de la precisión y la sensibilidad y es especialmente útil en situaciones donde se requiere un balance entre la precisión y el recall.

La variación en la tasa de aprendizaje y el tamaño de batch produjo F1-scores que oscilan entre 0.79 y 0.89, lo que indica una variabilidad considerable en el rendimiento dependiendo de la configuración. La inclusión de la regularización L1 y L2 también mostró un impacto notable, sugiriendo que una cuidadosa calibración de la regularización puede mejorar la capacidad del modelo para generalizar a datos no vistos.

El modelo con los parámetros por defecto (learning\_rate=0.1, batch\_size=128, regularization=0.0) muestra resultados prometedores con una convergencia estable de la pérdida y una mejora continua de la precisión a lo largo de las épocas de entrenamiento. En la Ilustración 2 podemos ver como las curvas de aprendizaje indican que el modelo no sufre de sobreajuste significativo, como lo demuestra la estrecha correspondencia entre el rendimiento en los conjuntos de entrenamiento y validación.

Imagen que contiene competencia de atletismo

Descripción generada automáticamente

Ilustración - Curvas de aprendizaje de modelo por defecto LeNet para Fashion-MNIST

Por otro lado, la matriz de confusión de la Ilustración 3 sugiere una clasificación sólida en la mayoría de las clases, aunque hay algunas confusiones, como entre las categorías 'Shirt' y 'T-shirt/top', que podrían atribuirse a similitudes visuales entre estas prendas.

Gráfico

Descripción generada automáticamente

Ilustración - Matriz de confusión de modelo por defecto LeNet para Fashion-MNIST

La visualización de los pesos de las capas y los mapas de características revela la capacidad del modelo para capturar características distintivas en las imágenes, lo que contribuye a su buen rendimiento. Las distribuciones de los pesos y sesgos en las diferentes capas sugieren que el modelo ha aprendido una variedad de patrones y no está sesgado hacia características particulares o específicas. Podemos apreciar esta información en la Ilustración 4.

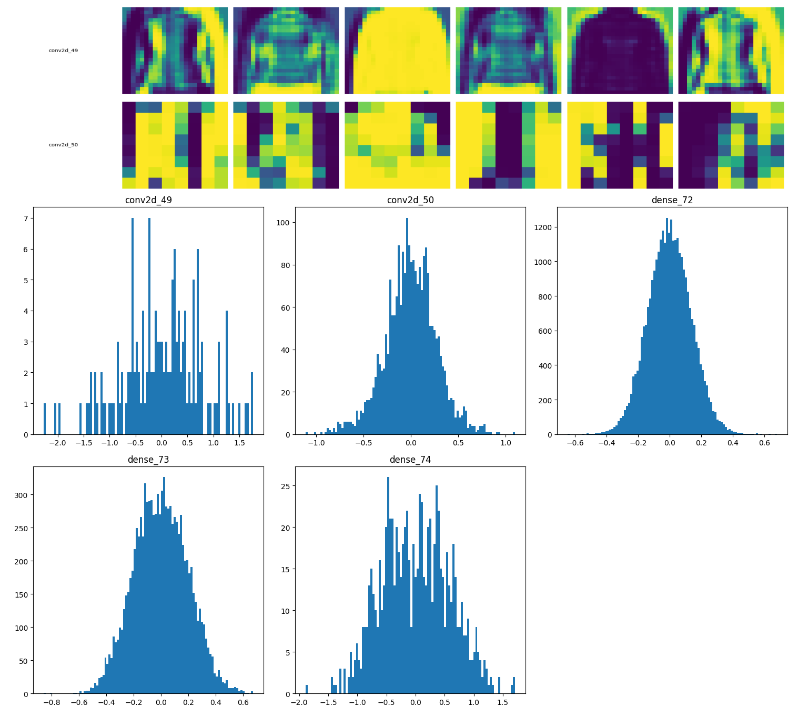


Ilustración - Pesos de las capas y mapas de características del modelo por defecto LeNet para Fashion-MNIST

Los resultados del modelo LeNet con valores por defecto demuestran que la arquitectura, incluso sin una optimización hiperparamétrica avanzada, es capaz de alcanzar un rendimiento decente en el conjunto de datos Fashion-MNIST. Esto refuerza la posición de LeNet como una arquitectura eficiente y efectiva para tareas de clasificación de imágenes a pequeña escala.

En segundo lugar, se escogió para un análisis más profundo el modelo que tiene una configuración de tasa de aprendizaje de 0.05, un tamaño de batch de 128 y sin regularización adicional. Este modelo fue seleccionado debido a su rendimiento superior en términos de métricas de evaluación, en su F1-score mayor frente al resto.

Este modelo muestra una convergencia de la pérdida y la precisión que es consistente y estable a lo largo de las épocas, similar al modelo por defecto. Sin embargo, lo notable es la mejora en la precisión de la validación que se mantiene cercana a la precisión de entrenamiento, lo que sugiere un buen equilibrio entre el ajuste y la generalización. Esto se puede apreciar en la Ilustración 5.

Imagen que contiene competencia de atletismo

Descripción generada automáticamente

Ilustración - Curvas de aprendizaje de modelo con mejores parámetros LeNet para Fashion-MNIST

La matriz de confusión de la Ilustración 6 revela un patrón de clasificación más preciso con menos errores entre clases confusas. Por ejemplo, la clasificación de 'Shirt' y 'T-shirt/top' muestra una mejora significativa, lo que indica una mejor capacidad de diferenciación entre categorías similares.

Gráfico

Descripción generada automáticamente

Ilustración - Matriz de confusión de modelo con mejores parámetros LeNet para Fashion-MNIST

Las visualizaciones de los pesos y los mapas de características ofrecen resultados parecidos a los obtenidos anteriormente, con lo cuál podemos extrapolar las mismas conclusiones que nos ha brindado el modelo con parámetros por defecto.

Comparando los dos modelos, este modelo parece ofrecer una mejora sustancial sobre el modelo por defecto, especialmente en términos de equilibrio entre el rendimiento de entrenamiento y de validación. La ausencia de regularización L1 o L2 en el mejor modelo sugiere que, para este conjunto de datos y arquitectura en particular, la complejidad del modelo no introduce un sobreajuste significativo y, por lo tanto, la regularización adicional no es necesaria para mantener la generalización.

Los resultados apoyan la importancia de una tasa de aprendizaje y un tamaño de batch bien ajustados para el éxito del entrenamiento de redes neuronales convolucionales. Además, confirman que un proceso meticuloso de ajuste de hiperparámetros puede conducir a mejoras significativas en la precisión de la clasificación sin necesidad de complicar el modelo con regularizaciones adicionales.

## AlexNet

Tras la evaluación de modelos basados en la arquitectura LeNet, se procedió a experimentar con la arquitectura AlexNet, una red neuronal convolucional más profunda y compleja, adaptada al conjunto de datos Fashion-MNIST. Siguiendo una metodología similar a la utilizada con LeNet, se exploraron diversas configuraciones de AlexNet para identificar aquella con el mejor rendimiento. Debido a las restricciones de recursos de la plataforma Google Colab, se limitó el análisis en profundidad a un solo modelo: el mejor obtenido según las métricas de rendimiento.

Para el entrenamiento de los modelos AlexNet, se utilizó una función que permitía ajustar dinámicamente la cantidad de filtros y neuronas a través de un factor de reducción, adaptando así la arquitectura original, que está diseñada para imágenes de mayor resolución, a las dimensiones de 28x28 píxeles de Fashion-MNIST.

* Varias capas convolucionales con una cantidad reducida de filtros, ajustados por el factor de reducción.
* Capas de pooling intercaladas para reducir la dimensionalidad y capturar las características dominantes.
* Dos capas densas (fully-connected) con un alto número de unidades, también ajustadas por el factor de reducción.
* Capas de Dropout para mitigar el sobreajuste mediante la regularización por omisión aleatoria de neuronas durante el entrenamiento.

El optimizador Adam se utilizó debido a su eficiencia y adaptabilidad en el ajuste de tasas de aprendizaje. La función de pérdida de entropía cruzada categórica con logits se seleccionó por su idoneidad en la clasificación multiclase. Un callback de parada temprana se incorporó para prevenir el sobreentrenamiento, deteniendo el entrenamiento si la precisión de validación no mejoraba tras un número especificado de épocas.

Se experimentó con variaciones en el factor de reducción, la tasa de aprendizaje y el tamaño del batch. Estos hiperparámetros se iteraron para entrenar múltiples modelos AlexNet, con cada uno siendo almacenado para análisis posteriores. Este proceso sistemático de iteración y almacenamiento aseguró que cada configuración pudiera ser evaluada y comparada meticulosamente.

### Resultados

Los experimentos con la arquitectura AlexNet han dado lugar a un conjunto de resultados que reflejan las capacidades y las limitaciones de esta red en el conjunto de datos Fashion-MNIST. El modelo de AlexNet, con una tasa de aprendizaje de 0.001 y un tamaño de batch de 128, muestra un rendimiento destacable según la métrica F1-score, que es indicativa de un equilibrio adecuado entre precisión y sensibilidad.

Al comparar con los resultados obtenidos del modelo LeNet, se observa que AlexNet, con su arquitectura más profunda, parece capturar mejor la complejidad de los datos de Fashion-MNIST. Esto se evidencia en los F1-scores superiores y en una matriz de confusión que muestra una mayor cantidad de predicciones correctas en casi todas las clases, especialmente en aquellas donde LeNet presentaba mayores confusiones, como entre 'Shirt' y 'T-shirt/top' como podemos ver en la Ilustración 7.

Gráfico

Descripción generada automáticamente

Ilustración - Matriz de confusión de modelo AlexNet para Fashion-MNIST

Las curvas de aprendizaje de la Ilustración 8 muestran una convergencia más rápida en las primeras épocas comparado con LeNet, lo que podría indicar una capacidad más efectiva de la red para adaptarse a las características del conjunto de datos desde el principio del entrenamiento. La precisión de validación permanece estrechamente alineada con la precisión de entrenamiento, lo cual es una buena señal de que el modelo generaliza bien y no está sobreajustado.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Ilustración - Curvas de aprendizaje de modelo AlexNet para Fashion-MNIST

Las visualizaciones de los pesos y los mapas de características muestran patrones de activación más complejos en AlexNet en comparación con LeNet, lo que sugiere que la red está aprendiendo representaciones más ricas y diversas de los datos. Esto es coherente con la arquitectura más profunda de AlexNet, que se diseñó para capturar jerarquías más complejas de características visuales.

Los resultados del experimento con AlexNet sugieren que, para el conjunto de datos Fashion-MNIST, una red más profunda y compleja puede proporcionar un rendimiento superior al de una red más simple como LeNet. La arquitectura avanzada de AlexNet permite una extracción de características más detallada y una mejor clasificación, lo que se traduce en un rendimiento mejorado. A pesar de las restricciones computacionales impuestas por Google Colab, el modelo de AlexNet con los mejores parámetros demuestra ser una elección robusta para la clasificación de imágenes de moda, resaltando la importancia de seleccionar una arquitectura y hiperparámetros adecuados para la tarea específica de clasificación de imágenes.

# Fine-tuning sobre modelos pre-entrenados

El avance hacia metodologías más sofisticadas en la clasificación de imágenes nos lleva al concepto de fine-tuning, una técnica en la que se ajustan modelos pre-entrenados a nuevos conjuntos de datos. Este enfoque aprovecha el conocimiento que los modelos han adquirido en tareas previas, lo cual es especialmente útil cuando se dispone de una cantidad limitada de datos o se desea ahorrar tiempo y recursos computacionales en el entrenamiento.

Para explorar la eficacia del fine-tuning, se ha elegido el corpus CIFAR-10, un conjunto de datos compuesto por 60,000 imágenes en color de 10 clases diferentes, que presenta un desafío mayor en comparación con conjuntos como Fashion-MNIST debido a su mayor variabilidad y complejidad de las imágenes.

ResNet, una arquitectura de red neuronal profunda conocida por su capacidad para entrenar redes con un número muy alto de capas es la elegida para este propósito. Su diseño incorpora conexiones residuales que permiten que las señales de gradiente fluyan a través de muchas capas sin atenuarse, lo que facilita el entrenamiento de redes muy profundas.

El proceso de fine-tuning ajustará un modelo ResNet pre-entrenado al corpus CIFAR-10, refinando sus pesos y sesgos para adaptarse a las especificidades de este nuevo conjunto de datos. Este enfoque permite evaluar cómo los conocimientos previos de un modelo sobre imágenes pueden transferirse y adaptarse a nuevas categorías de clasificación, una práctica que se está convirtiendo en un estándar en la visión por computadora debido a su eficiencia y efectividad.

## Corpus CIFAR-10

El conjunto de datos CIFAR-10, compuesto por 60,000 imágenes de color con dimensiones 32x32 repartidas en 10 clases distintas, representa un reto considerable en la clasificación de imágenes debido a su diversidad y complejidad visual. Las clases incluyen objetos y animales variados como aviones, coches, pájaros, gatos, ciervos, perros, ranas, caballos, barcos y camiones, cada una con 6,000 imágenes.

Para la preparación del conjunto de datos CIFAR-10, se ha seguido un procedimiento idéntico al aplicado en el corpus Fashion-MNIST donde se han cargado los datos, divido estos en entrenamiento, validación y prueba, asignando el 80% de las imágenes al entrenamiento y el 20% restante entre validación y prueba. También se ha realizado para que los valores de los píxeles estén en el rango [0, 1], mejorando la eficiencia del entrenamiento de la red así como un análisis estadístico para determinar la distribución del conjunto de datos, encontrando 40,000 imágenes para entrenamiento, 10,000 para validación y 10,000 para prueba.

## ResNet

En la serie de experimentos realizados con ResNet en el corpus CIFAR-10, se siguió un enfoque de fine-tuning, partiendo de un modelo ResNet50 pre-entrenado en ImageNet, una base de datos considerablemente más amplia y diversa. Este modelo base fue adaptado a las características del CIFAR-10 mediante el ajuste de las últimas capas a las 10 clases específicas del conjunto de datos. El proceso consistió en dos fases principales:

1. Creación del Modelo Base: Se obtuvo un modelo ResNet50 pre-entrenado, se eliminó su capa superior y se añadieron nuevas capas densas adaptadas al número de clases de CIFAR-10. Este modelo se denomina "vanilla" por ser la versión inicial sin ajustes específicos para el nuevo conjunto de datos.
2. Fine-Tuning: Se congeló el aprendizaje de todas las capas excepto las últimas cinco, permitiendo que solo estas capas ajusten sus pesos durante el entrenamiento en CIFAR-10. Este enfoque busca refinar los patrones de alto nivel aprendidos en ImageNet para que sean pertinentes para las clases de CIFAR-10.

El modelo "vanilla" ResNet muestra un F1-score inicial que sirve como punto de referencia para evaluar la eficacia del fine-tuning. La matriz de confusión y las visualizaciones de los mapas de características para este modelo reflejan el rendimiento antes de cualquier ajuste específico.

El modelo fine-tuned muestra una mejora en el F1-score, lo cual indica que el proceso de ajuste fino ha tenido un impacto positivo en la capacidad del modelo para clasificar correctamente las imágenes de CIFAR-10. Sin embargo, la matriz de confusión revela que aún existen desafíos en la clasificación de ciertas clases, lo que sugiere que puede haber un margen para una mayor optimización.

La mejora en el rendimiento del modelo fine-tuned confirma que los modelos pre-entrenados pueden ser efectivamente adaptados a nuevos conjuntos de datos con un enfoque de fine-tuning. Los resultados resaltan la potencia de transferir conocimientos de un modelo entrenado en un dominio más amplio y aplicarlo a un dominio más específico, lo cual es una estrategia de entrenamiento eficiente y eficaz.

A pesar de las mejoras observadas, la persistencia de algunas confusiones entre clases similares indica que aún hay espacio para la mejora del modelo. Puede ser beneficioso explorar un fine-tuning más granular o experimentar con diferentes técnicas de regularización y ajuste de hiperparámetros para maximizar la precisión de clasificación del modelo ResNet en CIFAR-10.

# Conclusiones

A lo largo de este trabajo, se ha llevado a cabo una exploración comprensiva de técnicas de aprendizaje profundo aplicadas a la clasificación de imágenes. Se han cubierto desde los enfoques más básicos y fundamentales con redes como LeNet hasta métodos avanzados que implican el fine-tuning de arquitecturas profundas como ResNet. Las conclusiones clave que se extraen de esta investigación son múltiples y variadas.

Experimentar con arquitecturas LeNet y AlexNet ha demostrado la importancia de los hiperparámetros y su optimización. Los resultados subrayan cómo pequeñas variaciones en estos pueden tener un impacto significativo en el rendimiento de los modelos. La experimentación ha validado la hipótesis de que redes más profundas y complejas, como AlexNet, son capaces de capturar características más abstractas y realizar clasificaciones más precisas, incluso en conjuntos de datos con imágenes de baja resolución.

El trabajo con ResNet ha demostrado que el fine-tuning es una técnica poderosa para adaptar conocimientos previos a nuevos conjuntos de datos. Especialmente en situaciones donde se cuenta con limitaciones de datos o de recursos computacionales, el fine-tuning se presenta como una alternativa eficiente para alcanzar un alto rendimiento sin necesidad de entrenar una red desde cero.

En definitiva, este trabajo ha proporcionado una visión valiosa de las técnicas modernas en clasificación de imágenes y ha subrayado la importancia de una metodología de experimentación sólida y reflexiva en el campo del aprendizaje profundo. Las lecciones aprendidas aquí pueden aplicarse a una amplia gama de aplicaciones prácticas en visión por computadora y más allá.