Práctica Final: Autoencoders determinísticos

Alejandro Díaz Cuéllar, Ilias El Hanouch Toudghi, Diego Navas Sánchez

Redes de Neuronas
Máster en Inteligencia Artificial Aplicada
2025-2026

1. Implementación

El objetivo principal fue comparar sistemáticamente el rendimiento de distintas arquitecturas de autoencoders en su capacidad para comprimir y reconstruir imágenes.

1.1. Arquitecturas evaluadas

Se evaluaron tres familias de modelos, identificando la mejor configuración de cada una:

- **Lineal (Baseline):** compuesto exclusivamente por capas `Linear`.
  Mejor config.: **'Linear_Base_3L_C30'** ($n\_layers = 3$, $C = 30$).
- **Convolucional (ConvAE):** utiliza capas `Conv2d` y `ConvTranspose2d`.
  Mejor config.: **'Conv_Base_3L_C100'** ($C = 100$).
- **Residual (ResNetAE):** arquitectura convolucional avanzada con bloques residuales (*skip connections*) para facilitar el entrenamiento profundo.
  Mejor config.: **'Residual_Large_128ch'** ($latent\_channels = 128$, $latent\_spatial = 7$).

1.2. Configuración de entrenamiento

Todos los modelos fueron entrenados con los siguientes parametros principales condiciones controladas para asegurar una comparativa equitativa:

- **Optimizador:** Adam.
- **Función de Pérdida:** Mean Squared Error (MSE).
- **Métrica de Evaluación:** Peak Signal-to-Noise Ratio (PSNR) en decibelios (dB).
- **Regularización Clave:** se implementó `EarlyStopping` para detener el entrenamiento cuando la pérdida de validación dejaba de mejorar, siendo esta la técnica de regularización más efectiva.

2. Resultados y configuraciones óptimas

La evaluación se centró en el PSNR máximo alcanzado en el conjunto de test. La arquitectura `Residual_Large_128ch` demostró ser la mejor configuración de forma indiscutible en ambos datasets.

La siguiente tabla 2 resume los mejores resultados obtenidos para cada familia de modelos, utilizando los datos de las ejecuciones del notebook.

Mejor configuración por tipo	Dataset	Dim. Latente (z)	PSNR (dB)
'Linear_Base_3L_C30'	MNIST	30	20.63 dB
'Linear_Base_3L_C30'	F-MNIST	30	19.51 dB
'Conv_Base_3L_C100'	MNIST	100	27.89 dB
'Conv_Base_3L_C100'	F-MNIST	100	25.81 dB
'Residual_Large_128ch'	**MNIST**	**128**	**40.35 dB**
'Residual_Large_128ch'	**F-MNIST**	**128**	**34.88 dB**

2.1. Análisis de regularización

Se experimentó con regularización L1 (sparse), L2 (weight decay) y Dropout.

- **Impacto Negativo:** las técnicas de regularización explícitas (L1, L2, Dropout) degradaron el rendimiento de reconstrucción (PSNR) en comparación con los modelos base.
- **Ejemplo:** el modelo 'Residual_96ch_+_L2' (36.95 dB) rindió significativamente peor que el 'Residual_Base_64ch' (39.67 dB) o el 'Residual_Large_128ch' (40.35 dB) en MNIST.

La conclusión principal es que para la tarea de reconstrucción pura, el mecanismo de `EarlyStopping` (con paciencia de 8-10 épocas) fue el regularizador más efectivo y suficiente, previniendo el sobreajuste sin sacrificar fidelidad. En los modelos residuales como el mostrado en la propia tabla para MNIST, de hecho se agotaron las 80 épocas límites máxima, pudiéndo haberse extendido incluso más pero se ha dejado así por razones de límites de tiempo y cómputo.

3. Eliminación de ruido (Denoising Autoencoder)

Se entrenó el enfoque residual anterior para reconstruir imágenes limpias a partir de versiones con ruido gaussiano añadido a través de los mejores modelos. Los resultados (PSNR) variaron según la intensidad del ruido, como se muestra en la Tabla 3.

Dataset	Varianza de Ruido	PSNR (dB)
MNIST	0.01	30.61 dB
MNIST	0.04	26.76 dB
MNIST	0.09	24.47 dB
F-MNIST	0.04	24.08 dB

El rendimiento base es bastante notable (30.61 dB), pero se nota, como era de esperar que a mayor varianza de ruido, mayores dificultades para reconstruir.

4. Análisis de pruebas adicionales

Se realizaron cuatro experimentos "extra"para profundizar en las capacidades de los autoencoders, utilizando la arquitectura base residual como la mejor entre las evaluadas.

4.1. Diferentes tamaños del dataset

Se re-entrenaron las mejores configuraciones de cada familia (Lineal, Convolucional y Residual) usando subconjuntos de los datos de entrenamiento de MNIST (10 %, 20 %, 50 % y 100 %) para evaluar su eficiencia y robustez:

Tamaño del dataset MNIST	Lineal (PSNR)	Conv (PSNR)	Residual (PSNR)
25 %	13.19 dB	23.17 dB	36.67 dB
50 %	13.32 dB	23.66 dB	38.83 dB
75 %	13.37 dB	23.43 dB	39.54 dB
100 %	13.37 dB	24.14 dB	39.89 dB

Y del mismo modo en FMNIST:

Tamaño del dataset FMNIST	Lineal (PSNR)	Conv (PSNR)	Residual (PSNR)
25 %	14.40 dB	21.51 dB	32.67 dB
50 %	14.45 dB	21.86 dB	33.60 dB
75 %	14.28 dB	20.61 dB	34.04 dB
100 %	14.53 dB	21.58 dB	34.58 dB

La arquitectura Residual ('Residual_Large_128ch') no solo es superior en general, sino que es drásticamente más eficiente con los datos. Con solo el 10 % de los datos de entrenamiento (~38.0 dB), el modelo residual supera por un margen enorme a los modelos Lineal y Convolucional entrenados con el 100 % de los datos (20.63 dB y 27.89 dB, respectivamente). Aparte de esto, se demuestra que el tamaño del dataset influye en los resultados generales, ya que con más datos, mejor rendimiento.

4.2. Pruebas cruzadas con los dataset

Se evaluó la capacidad de generalización del mejor modelo entrenado en MNIST, probándolo directamente sobre el conjunto de test de F-MNIST.

Modelo Entrenado	Dataset de Test	PSNR (dB)
En MNIST	F-MNIST	23.72 dB
En F-MNIST (Referencia)	F-MNIST	34.88 dB

El rendimiento del modelo entrenado en MNIST al ser probado en F-MNIST (23.72 dB) es significativamente inferior (una diferencia de 11.16 dB) comparado con el modelo nativo de F-MNIST. Esto confirma que los autoencoders aprenden representaciones altamente específicas del dominio de los dígitos en el que fueron entrenados, y el espacio latente de dígitos no generaliza bien al dominio de la ropa.

4.3. Análisis del espacio latente

Se utilizaron PCA y t-SNE para visualizar las representaciones latentes generadas por el ResNetAE.

- **MNIST:** mostró una separabilidad de clases excelente. Los dígitos formaron *clusters* densos y bien definidos.
- **F-MNIST:** mostró una estructura más compleja y con mayor solapamiento entre clases (ej. "T-shirt", "Pullover", "Coat").

Este análisis reveló que MNIST permite representaciones más compactas, mientras que F-MNIST, al ser más complejo, se beneficia de espacios latentes de mayor dimensionalidad (como el de 128 canales) para capturar su variabilidad.

5. Conclusiones

El enfoque residual, concretamente el de los 128 canales de dimensionalidad, fue la arquitectura superior, alcanzando un PSNR muy elevado en MNIST y 34.88 dB en F-MNIST. Las conexiones residuales demostraron ser críticas para entrenar redes profundas capaces de preservar detalles finos.

Se concluye que las técnicas de regularización explícitas (L1, L2, Dropout) no han resultado del todo eficaces para este dominio de reconstrucción pura, siendo el `EarlyStopping` la estrategia dominante.

Finalmente, los análisis adicionales confirmaron que nuestros autoencoders son efectivos para la eliminación de ruido, pero aprenden representaciones altamente específicas de su dominio de entrenamiento, como evidenció el fallo total en las pruebas de transferencia de conocimiento.