

Classificação de imagens utilizando redes neurais convolucionais

Leonardo Loureiro Costa
leonardo.costa@unifesp.br

Resumo—Redes neurais convolucionais são vastamente utilizadas para classificação de imagens. Neste relatório exploraremos o uso desse modelo neural na classificação de dígitos do dataset MNIST.

Mostraremos ser possível adquirir uma acurácia superior a 99,3% na classificação de dígitos, e que é possível obter resultados melhores e de formas mais rápidas nessa tarefa do que usando redes MLP.

Exploraremos como ruídos gaussianos afetam no desempenho de cada uma das redes e como um filtro passa-baixa altera esse resultado.

I. INTRODUÇÃO

Redes neurais convolucionais são estado-da-arte em tarefas de classificação de imagens. Os filtros aplicados por meio de convolução dão a esse tipo de rede a capacidade de melhor compreender padrões, mesmo quando esses padrões sofrem escala, rotação ou translação.

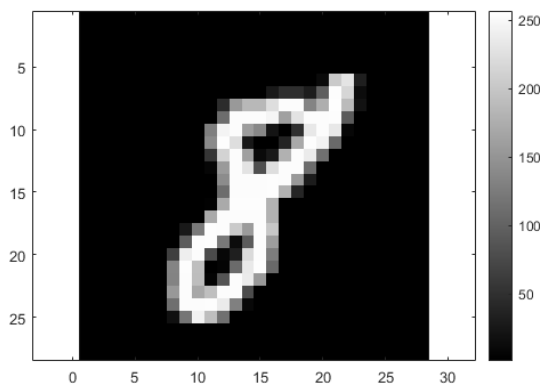


Figura 1: Exemplo de imagem do dataset MNIST

Neste relatório será apresentado como uma rede neural convolucional pode ser utilizada para a classificação de dígitos em imagens do conjunto MNIST.

Será comparado, também, o desempenho dessa rede convolucional com uma simples rede MLP.

O dataset MNIST consiste em 70 mil imagens de canal único, de profundidade 8 bits, 28 por 28 píxeis de algarismos de 0 a 9 em brancos representados em um fundo preto.

Dessas 70 mil imagens, 54 mil serão usadas para treino, 6 mil para validação e 10 mil para teste.

II. OBJETIVOS

- Avaliar diferentes topologias no aprendizado do modelo
- Calcular a matriz de confusão do melhor modelo
- Avaliar a diferença entre os dois melhores modelos
- Comparar o desempenho da melhor rede convolucional com o da melhor rede MLP.

III. RESULTADOS

A. Modelos convolucionais

Os modelos treinados possuem a topologia ilustrada na figura 2. O número de neurônios por camada de convolução de cada modelo está como mostra a tabela I.

Todas as redes foram treinadas em 20 épocas de lote tamanho 32. O modelo 5 é idêntico ao modelo 4, porém não possui a penúltima camada — Dropout.

Modelo	Nº neurônios (Conv2D 1 e Conv2D 2)
Modelo 1	16 e 8
Modelo 2	64 e 32
Modelo 3	128 e 64
Modelo 4	512 e 256
Modelo 5	512 e 256

Tabela I: Número de neurônios por camada de convolução

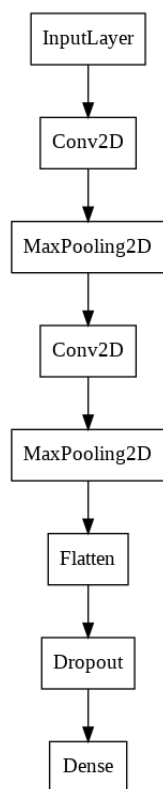


Figura 2: Modelo utilizado.

As 5 redes foram treinadas durante as épocas especificadas e apresentam a evolução das acurácias como ilustra a tabela 3. Nota-se que os modelos 3 e 4 obtiveram os melhores desempenhos.

A figura 4 ilustra as curvas da figura 3, porém com um filtro de média móvel de janela 5 aplicado. Dessa forma é mais fácil visualizar que o modelo que melhor aprendeu, e que melhor performou, é a rede 4.

Observa-se, também, que o Modelo 5 apresentou o segundo pior resultado dos testes realizados.

1) Comparando os 2 melhores modelos convolucionais: Os modelos 3 e 4 obtiveram os melhores resultados na classificação de dígitos. Ambos foram treinados novamente, porém por 50 épocas.

As figuras 5 e 6 ilustram o desempenho de cada um dos modelos na acurácia do conjunto de validação durante o treino, sem e com um filtro de média móvel de janela 5, respectivamente.

Observa-se que o desempenho de ambos é muito similar, a diferença é mínima.

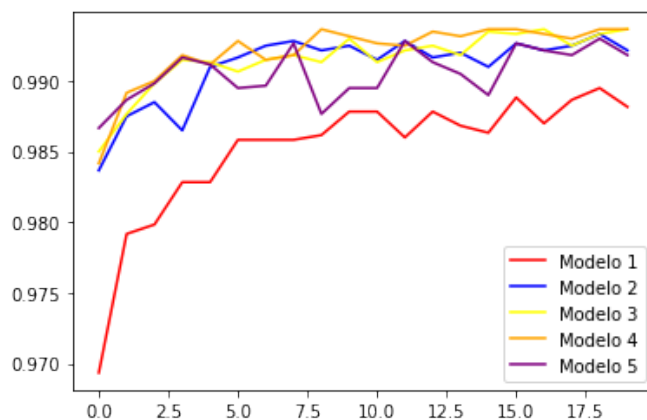


Figura 3: Acurácia no conjunto de validação dos 5 modelos apresentados.

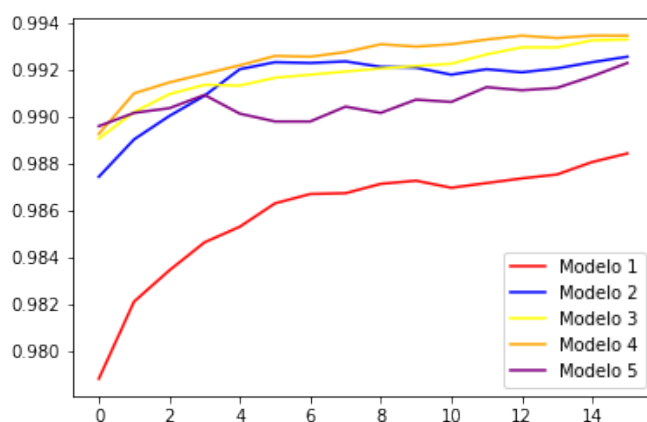


Figura 4: Acurácia no conjunto de validação dos 5 modelos apresentados com um filtro de média móvel com janela 5.

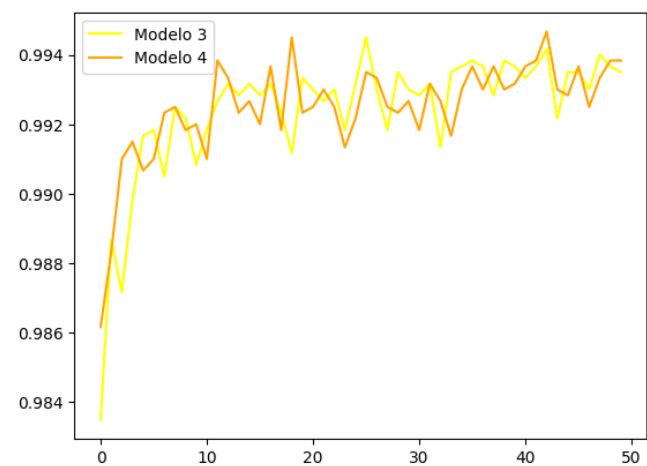


Figura 5: Acurácia no conjunto de validação dos modelos 3 e 4.

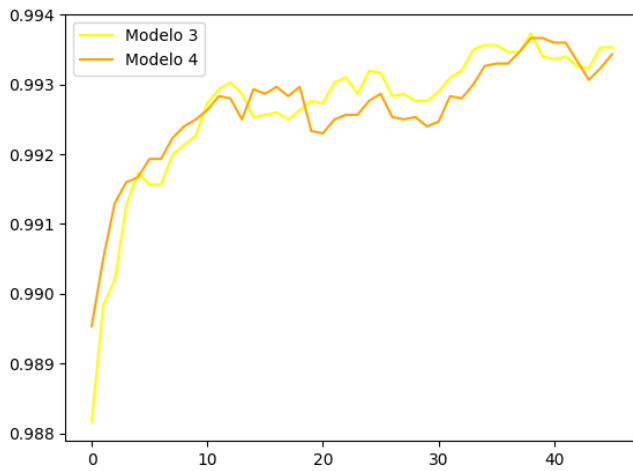


Figura 6: Acurácia no conjunto de validação dos modelos 3 e 4 com um filtro de média móvel de janela 5.

B. Modelo MLP

A rede MLP foi treinada utilizando a mesma base de dados, obteve uma acurácia de $\approx 98\%$ no conjunto de testes.

Os parâmetros dessa rede estão presentes na tabela II

Camada	Nº Perceptrons	Ativação
1ª camada	1024	ReLu
2ª camada	1024	ReLu e Regularização L2
3ª camada	10	SoftMax

Tabela II

C. Comparação entre CNN e redes MLP

A fim de comparar o desempenho de uma CNN com uma rede MLP, a rede MLP em questão foi comparada com o modelo 4 da CNN proposta nesse relatório, a que obteve melhores resultados.

É trivial que a CNN possui uma acurácia maior na tarefa de classificação no dataset proposto — MNIST —, cerca de 1 ponto percentual a mais. As figuras 7 e 8 ilustram respectivamente a matriz de confusão das redes MLP e CNN.

As linhas das tabelas indicam o valor real, verdadeiro, das imagens; as colunas, o valor predizado pelos modelos neurais.

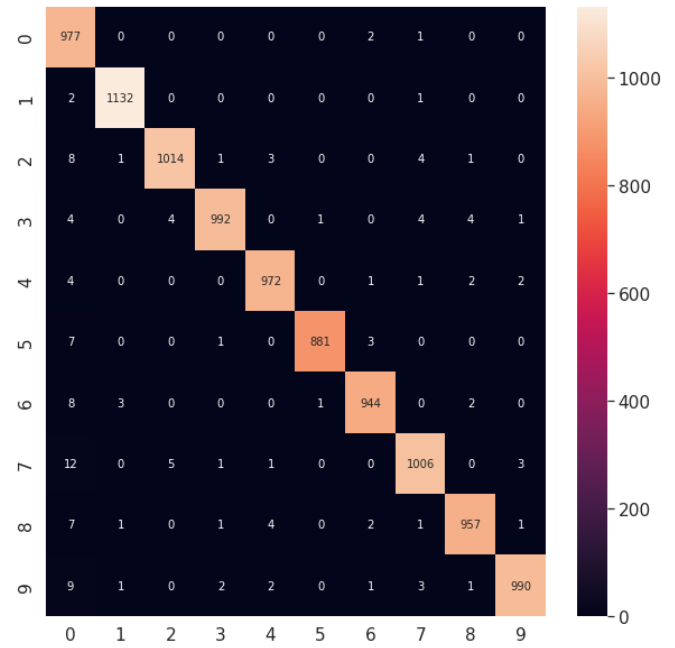


Figura 7

Observa-se que a rede CNN obteve menos erros, classificou mais corretamente as imagens em respeito aos seus verdadeiros valores.

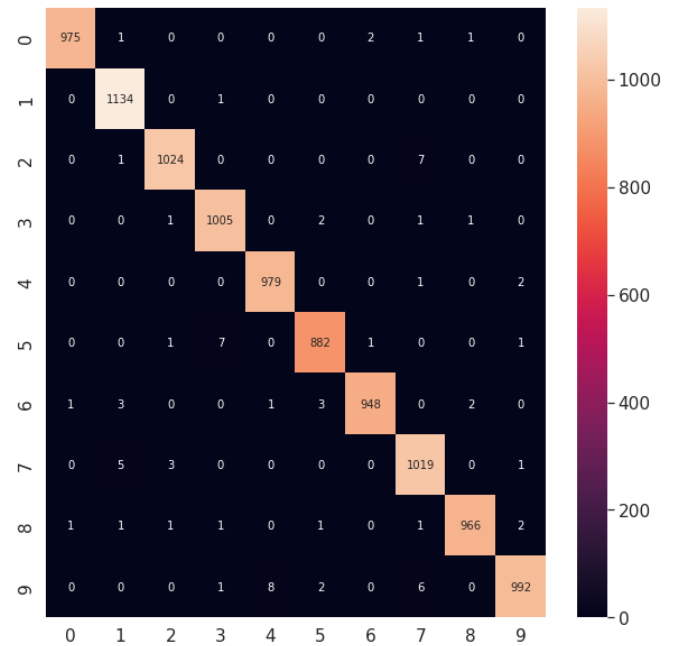


Figura 8

Essa vantagem de redes CNN, porém, não se mantém quando ruídos são introduzidos à imagem.

Para testar o desempenho de cada um dos modelos a base de dados do conjunto de testes passou por

2 processo de tratamento de imagens: a aplicação de um ruído gaussiano centrado em 0 com $\sigma = 0.1$, e a aplicação de um filtro passa-baixa de média sobre essa imagem de janela 2 por 2.

A figura 9 ilustra o dígito 9, um elemento do conjunto de testes, com ambos os processos de ruído e filtros aplicados.



Figura 9: Imagem original contendo o dígito 9, imagem com um ruído gaussiano aplicado, imagem com um filtro passa baixa aplicado sobre o ruído gaussiano.

Cada rede avaliou a sua acurácia nos conjuntos de testes com ruído e com o ruído filtrado, as figuras 10 e 11 ilustram, respectivamente, seus desempenhos.

O eixo das abscissas representa x , onde $\sigma = \frac{x}{20}$, dessa forma, quanto maior o valor de x , maior o valor de σ ; maior a intensidade do ruído.

Observa-se que na figura 10 a CNN obtém maiores valores de acurácia na classificação de dígitos até $\approx x = 8$, após esse valor, a rede MLP detém a liderança.

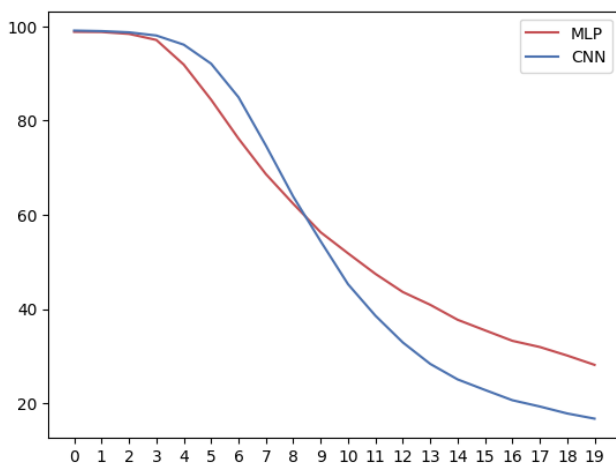


Figura 10: Acurácia da CNN e da rede MLP na classificação das imagens em função da intensidade do ruído gaussiano

Na figura 11 observa-se um comportamento diferente, a CNN detém a liderança na acurácia de classificação para todo $0 \leq x \leq 19$.

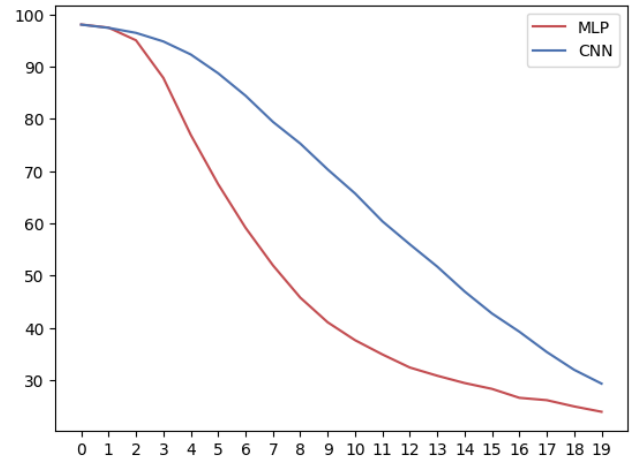


Figura 11: Acurácia da CNN e da rede MLP na classificação das imagens em função da intensidade do ruído gaussiano, com filtro passa-baixa aplicado.

IV. CONCLUSÃO

Por meio dos experimentos realizados exemplifica-se a superioridade de redes convolucionais em tarefas de classificação de imagens como as do dataset MNIST: requer menor tempo de treinamento, possui menor complexidade e oferece resultados com maior acurácia.

Redes convolucionais proporcionam tais feitos porque detém de um "pré-processamento" de filtros convolucionais que detectam padrões mais eficientemente do que redes MLP.

Observa-se que, na adição de ruídos como o gaussiano, a rede MLP ganha liderança em desempenho para valores de $\sigma > \approx 0.25$. Essa liderança perde lugar para as redes CNN quando o ruído é atenuado com filtro passa baixa de média: redes CNN se mostraram superiores em reconhecer padrões nessas imagens.

REFERÊNCIAS

- [1] Digitos mnist, 2022.
- [2] Simon S. Haykin. *Neural networks and learning machines*. Pearson Education, Upper Saddle River, NJ, third edition, 2009.