

Trabalho Bimestral

Bernardo Fagnani Carneiro

Leonardo da Silva Muller

Marco Antonio Pagno Geraldo

I. INTRODUÇÃO

A inteligência artificial está em constante evolução, impulsionada pelas arquiteturas avançadas de redes neurais. Entre elas, destacam-se a ResNet, a DenseNet e as arquiteturas de redes neurais convolucionais (CNN). A ResNet resolve o problema de degradação do desempenho em redes profundas usando conexões residuais. A DenseNet possui uma estrutura densamente conectada que melhora o fluxo de informações. As CNNs são fundamentais para o processamento de dados em formato de grade, como imagens, e permitem a extração de características relevantes. Essas arquiteturas têm alcançado resultados notáveis na visão computacional.

II. RESNET

A RESNET (Residual Neural Network) é uma arquitetura de rede neural profunda desenvolvida para superar o desafio de treinar redes muito profundas. Foi proposta por Kaiming He, Xiangyu Zhang, Shaoqing Ren e Jian Sun em 2015 e desde então se tornou uma das arquiteturas mais populares em tarefas de visão computacional.

Uma das principais dificuldades ao treinar redes neurais profundas é o problema do desvanecimento do gradiente, no qual os gradientes se tornam muito pequenos à medida que são propagados para as camadas iniciais da rede. Isso dificulta o aprendizado eficiente dessas camadas e pode levar a um desempenho inferior. Além disso, redes mais profundas também podem sofrer de problemas de degradação, onde adicionar camadas extras leva a uma degradação da precisão devido à dificuldade de otimização.

A. Arquitetura da RESNET

A RESNET possui várias variantes, como a RESNET-18, RESNET-34, RESNET-50, entre outras. O número após o nome indica o número de camadas na rede. Por exemplo, a RESNET-50 possui 50 camadas.

A arquitetura da RESNET geralmente começa com uma única camada convolucional seguida por uma camada de pooling para reduzir a resolução espacial da entrada. Em seguida, são empilhados vários blocos residuais, que consistem em camadas de convolução e conexões residuais. No final, há uma camada de média global ou uma camada totalmente conectada seguida por uma camada de saída, dependendo da tarefa em questão.

B. Benefícios e Aplicações

A RESNET trouxe benefícios significativos para várias tarefas de visão computacional. Sua capacidade de treinar redes profundas levou a um desempenho aprimorado em tarefas como classificação de imagens, detecção de objetos, segmentação semântica e muito mais.

C. Conclusão

A RESNET é uma arquitetura poderosa que supera os desafios de treinamento de redes profundas. Sua utilização de blocos residuais permite que as informações fluam diretamente através das camadas, facilitando o treinamento e melhorando o desempenho em várias tarefas de visão computacional. A RESNET continua a ser uma área ativa de pesquisa e uma escolha popular para a comunidade de aprendizado profundo.

III. DENSENET

O DenseNet (Densely Connected Convolutional Network) é uma arquitetura de rede neural profunda desenvolvida por Gao Huang, Zhuang Liu, Laurens van der Maaten e Kilian Weinberger em 2016. Essa arquitetura introduziu uma abordagem inovadora para conectar camadas em redes neurais convolucionais, resultando em uma representação densa e altamente conectada.

A. Estrutura da Rede DenseNet

A estrutura do DenseNet consiste em várias unidades chamadas de blocos Dense. Cada bloco Dense é composto por várias camadas convolucionais em sequência, seguidas por uma conexão densa. Essa conexão densa concatena todas as saídas das camadas anteriores e as alimenta como entrada para as camadas subsequentes.

B. Benefícios e Aplicações

O DenseNet possui várias vantagens significativas. Ao permitir que todas as camadas tenham acesso direto a informações anteriores, ele promove o reuso eficiente de recursos e melhora o gradiente Flow, facilitando o treinamento de redes mais profundas. Além disso, o DenseNet é mais compacto em termos de parâmetros em comparação com outras arquiteturas, o que o torna computacionalmente mais eficiente.

C. Conclusão

O DenseNet é uma arquitetura de rede neural poderosa que promove conexões densas entre camadas, permitindo fluxo de informações mais eficiente e reuso de recursos em redes neurais profundas. Sua estrutura densamente conectada e eficiência em termos de parâmetros tornaram o DenseNet uma escolha popular na área de visão computacional e continua sendo objeto de pesquisa ativa para melhorias e adaptações em várias aplicações.

IV. CNN

As Redes Neurais Convolucionais (Convolutional Neural Networks, CNN) são uma classe especializada de redes neurais profundas, projetadas especificamente para lidar com dados de natureza espacial, como imagens. Elas foram desenvolvidas com base na inspiração do córtex visual do cérebro humano e se tornaram uma das arquiteturas mais populares e bem-sucedidas em tarefas de visão computacional.

A. Estrutura e Funcionamento

As CNNs são redes neurais especializadas em processamento de imagens. Elas usam camadas convolucionais para extrair características importantes das imagens e camadas de pooling

para reduzir a dimensionalidade. Essas camadas são intercaladas, formando uma hierarquia de representações abstratas. Por fim, camadas totalmente conectadas são usadas para classificar ou fazer previsões com base nas características extraídas.

B. Benefícios e Aplicações

As CNNs têm vantagens como invariância a translações, reuso de parâmetros e hierarquia de representações, tornando-as adequadas para tarefas de visão computacional. Elas são amplamente utilizadas em classificação de imagens, detecção de objetos, segmentação semântica, reconhecimento facial e processamento de vídeo.

C. Conclusão

As CNNs são uma classe especializada de redes neurais profundas projetadas para tarefas de visão computacional. Com suas camadas convolucionais e de pooling, elas são capazes de extrair características relevantes das imagens e aprender representações hierárquicas, permitindo o reconhecimento de padrões complexos. As CNNs têm sido amplamente aplicadas em uma variedade de tarefas, desde classificação de imagens até processamento de vídeo, e continuam sendo uma área ativa de pesquisa e desenvolvimento na área de visão computacional.

V. SOLUÇÃO DO PROBLEMA

A. Problema

Criar um classificador de frutas, onde é fornecido vários tipos de frutas. Utilizando uma CNN, Resnet e Densenet criar 3 modelos para classificar as frutas.

B. Dataset

Utilizamos o Dataset Fruit – 262, com as seguintes classes: acai, apple, avocado, banana blueberry, carabola, coconut, coffee, durian, grapefruit, jaboticaba e kiwi. Sendo um total de 12745 imagens, onde utilizamos 10196 (80%) para treino, e e o restante para validação.

C. CNN

Inicialmente, a base de dados foi dividida, reservando 20% das imagens para fins de validação. Em seguida, as imagens foram redimensionadas para o tamanho de 32x32 pixels e o tamanho do lote (batch size) foi definido como 128.

O processo de definição das camadas da rede neural prosseguiu da seguinte forma: primeiramente, foi adicionada uma camada convolucional com 64 filtros e um tamanho de kernel de 3x3. Em seguida, foi aplicada uma camada de normalização. Posteriormente, outra camada convolucional com 64 filtros e um tamanho de kernel de 5x5 foi adicionada, seguida novamente por uma camada de normalização e uma camada de max pooling.

Essa sequência de camadas foi repetida, porém com 128 filtros. Após a última camada convolucional, o output foi achatado (flattened) e foram realizados testes com diferentes números de épocas. No final desses treinamentos, obteve-se uma acurácia de 72,97% após 15 épocas.

D. Resnet

Primeiramente, a base de dados foi dividida, reservando 20% das imagens para fins de validação. Em seguida, as imagens foram redimensionadas para o tamanho de 32x32 pixels, e um tamanho de lote (batch size) de 128 foi definido para o treinamento do modelo.

Ao iniciar a definição das camadas do modelo, foram utilizadas três camadas convolucionais com 64 filtros cada, seguidas de camadas de normalização entre elas. Posteriormente, foram adicionadas mais três camadas convolucionais, desta vez com 128 filtros, novamente com uma camada de normalização no meio. Por fim, foram acrescentadas mais três camadas convolucionais com 256 filtros, também com uma camada de normalização no meio.

Após as camadas convolucionais, o output da última camada foi achatado (flattened). Em seguida, foram realizados testes com diferentes números de épocas de treinamento. Ao final

desses experimentos, o modelo alcançou uma acurácia de 74,81%.

E. Densenet

Conforme já apresentado no Resnet, somente foi modificado alguns parâmetros para o Densenet:

Inicialmente, a base de dados foi dividida, destinando 20% das imagens para fins de validação. Em seguida, as imagens foram redimensionadas para o tamanho de 64x64 pixels, e o tamanho do lote (batch size) foi configurado como 128.

O modelo foi construído com várias camadas. No início, foram utilizadas três camadas sequenciais, cada uma com 64 filtros, seguidas por uma camada de normalização entre elas. Posteriormente, o mesmo padrão foi aplicado com 128 filtros, novamente com uma camada de normalização no meio, e, por fim, com 256 filtros, também acompanhados de uma camada de normalização intermediária.

Após a última camada, o output foi achatado, convertendo-o em um vetor unidimensional. Foram conduzidos testes com diferentes números de épocas durante o treinamento. Ao final, o modelo alcançou uma acurácia de 70,02%."

F. Tabela de comparação das acurácias

a) Foram feitos diversos testes em nossas redes neurais para obter valores de acurácias mais aprimoradas, conforme tabela:

TABELA

Comparação da acurácia		
Modelo	Épocas	Acurácia
Resnet	10	69,02%
Resnet	15	71,17%
Resnet	20	74,81%
Densenet	10	15,62%
Densenet	15	53,33%
Densenet	20	70,02%
CNN	10	71,40%
CNN	15	72,96%
CNN	20	66,97%

Conforme a tabela podemos observar que a CNN desenvolvida por nós obteve o melhor resultado com 10 épocas, já com 15 épocas a CNN novamente obteve a melhor performance e com 20 épocas a Resnet obteve melhor performance.

GRÁFICO DE FUNÇÕES DE PERDA

Um gráfico de "% por época" que compara a loss (perda) e a val_loss (perda de validação) é comumente utilizado para monitorar o desempenho de um modelo de aprendizado de máquina durante o treinamento. A perda é uma medida que indica quão longe as previsões do modelo estão dos valores reais durante o treinamento, enquanto a val_loss mede o desempenho do modelo em um conjunto de dados de validação separado, que não foi utilizado durante o treinamento.

Um caso curioso que aconteceu foi o do DenseNet em que a loss diminui significativamente, mas a val_loss começa a aumentar, é sinal de que o modelo está memorizando os dados de treinamento e não generalizando adequadamente para novos exemplos, resultando em overfitting.

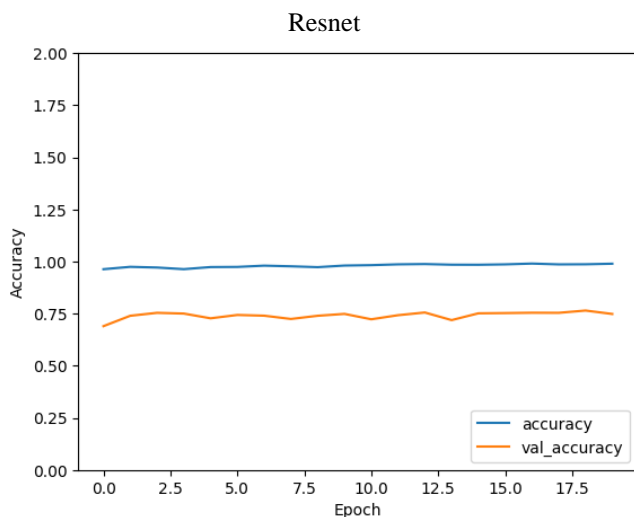
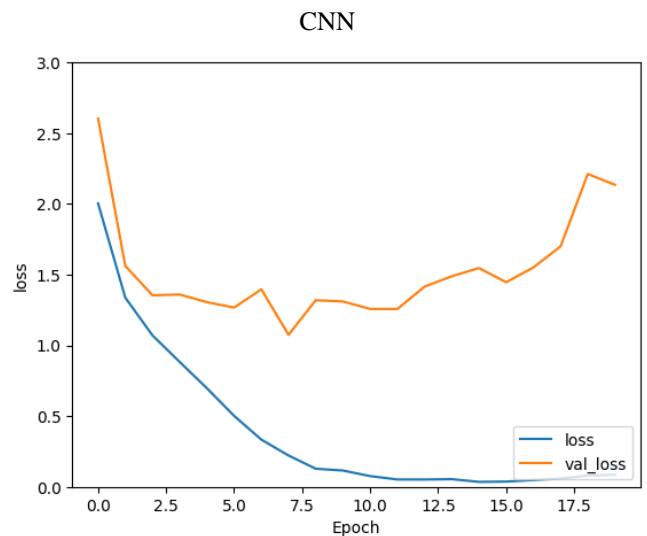
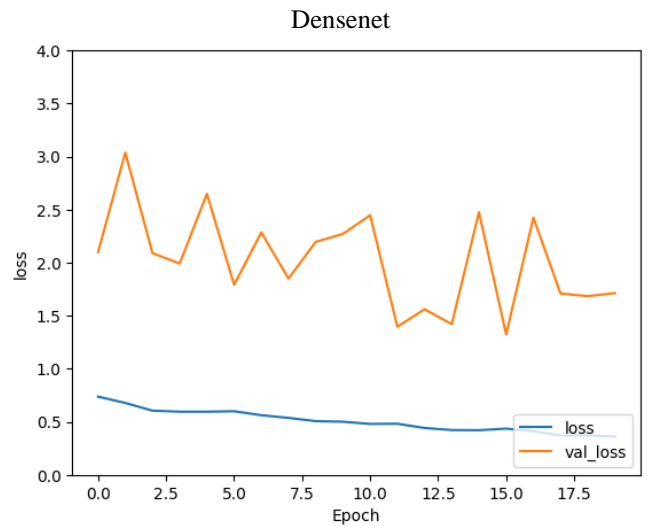


GRÁFICO DE ACURÁCIA POR ÉPOCA

Um gráfico de "% por época" que compara a accuracy (acurácia) e a val_accuracy (acurácia de validação) é comumente usado para avaliar o desempenho de um modelo de aprendizado de máquina durante o treinamento. A acurácia é uma medida que indica a proporção de previsões corretas do modelo em relação ao total de amostras, enquanto a val_accuracy mede o desempenho do modelo em um conjunto de dados de validação separado.

O mesmo caso do DenseNet pode ser novamente visualizado no DenseNet, a accuracy aumentou, mas a val_accuracy estagnou e diminuiu, é um sinal de que o modelo está superajustando (overfitting) os dados de treinamento e não está generalizando adequadamente para novos exemplos.

CONCLUSÃO

A análise comparativa entre as redes convolucionais CNN, DenseNet e ResNet teve como objetivo principal ensinar o uso dessas arquiteturas no processamento de imagens. O conjunto de dados utilizado consistia em 12 classes, com 1000 imagens cada, totalizando 12000 imagens disponíveis para treinamento e avaliação dos modelos. A divisão dos conjuntos foi feita de forma tradicional, com 80% das imagens destinadas ao treinamento e 20% para a validação.

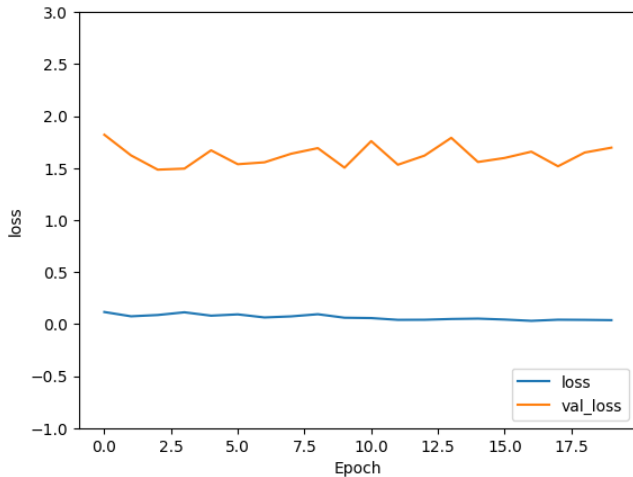
Durante o processo de treinamento e avaliação, foram obtidas as acurácias mais altas para cada uma das redes. A ResNet se destacou, alcançando uma acurácia de 74%, enquanto o DenseNet obteve 70% e uma CNN própria alcançou 73%. Esses resultados nos permitem concluir que a ResNet apresentou o melhor desempenho em termos de acurácia para esse conjunto de dados específico.

No entanto, um aspecto relevante que merece atenção é a possível ocorrência de overfitting no DenseNet. A arquitetura do DenseNet é conhecida por sua complexidade, o que pode resultar em um alto poder de aprendizado. No entanto, quando combinada com um conjunto de dados relativamente pequeno, como foi o caso deste estudo, a complexidade do DenseNet pode se tornar um fator de risco para o overfitting.

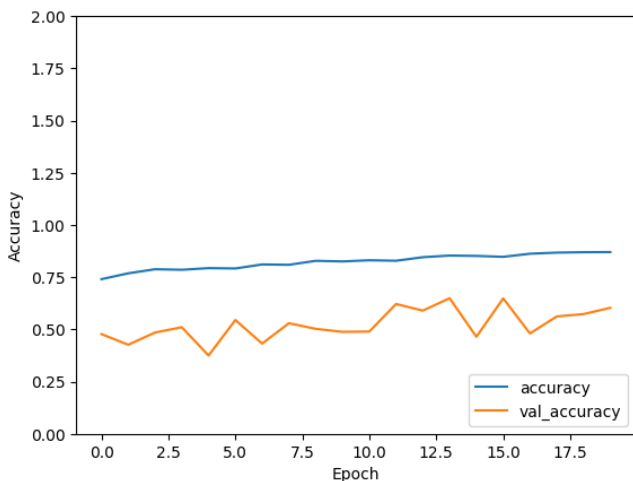
Indícios de overfitting no DenseNet foram observados por meio da análise comparativa entre as métricas de treinamento e validação. Foi observada uma divergência entre a acurácia e o erro de validação (val_accuracy e val_error) em relação às métricas de treinamento (accuracy e error). Essa divergência, caracterizada pela regressão e instabilidade no gráfico comparativo, indica a presença de overfitting. Esses sinais reforçam a hipótese de que a complexidade do DenseNet pode ter sido inadequada para o tamanho limitado do conjunto de dados utilizado.

Dessa forma, com base nos resultados obtidos, podemos concluir que a ResNet se destaca como a melhor escolha entre as três redes avaliadas para esse conjunto de dados específico. Além disso, os indícios de overfitting observados no DenseNet ressaltam a importância de selecionar uma arquitetura adequada ao tamanho e complexidade do conjunto

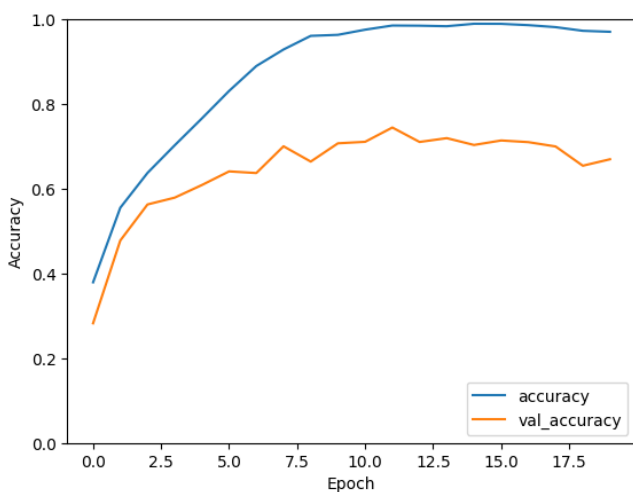
Resnet



Densenet



CNN



de dados disponível, a fim de evitar problemas de overfitting e alcançar melhores resultados de generalização.

Esse estudo demonstrou a relevância de compreender as características das diferentes arquiteturas de redes convolucionais e considerar cuidadosamente a seleção do

modelo mais apropriado para cada tarefa. Através desse trabalho, foi possível adquirir conhecimento sobre o funcionamento das CNNs e explorar suas aplicações no processamento de imagens, contribuindo para o avanço do campo de aprendizado de máquina e fornecendo diretrizes para futuras pesquisas nessa área.