

CNN vs DINO V2

Bruno Andrejewski Roglin

Birds 20 Species Image Classification Dataset

Visão Geral:

- Total de Imagens: 3.408
- Imagens de Treinamento: 3.208
- Imagens de Teste: 100 (5 por espécie)
- Imagens de Validação: 100 (5 por espécie)

Características:

- Dimensões da Imagem: 224 x 224 x 3 (RGB)
- Qualidade: Alta - Uma ave por imagem, cobrindo >50% da imagem

Birds 20 Species Image Classification Dataset

Estrutura de Dados:

- Diretórios: 475 (um por espécie de ave)
- Arquivo CSV: Contém caminhos dos arquivos, rótulos, nomes científicos, informações do conjunto e class_id

Pré-processamento:

- Sem Augmentação: Imagens originais
- Processamento: Verificação de duplicatas, recorte (>50% de cobertura), redimensionamento

Birds 20 Species Image Classification Dataset

Recomendação de Treinamento:

- Tamanho da Imagem: Sugerido 150 x 150 x 3 para treinamento mais rápido
- Equilíbrio: Conjunto de treinamento variado, mas pelo menos 130 imagens por espécie
- Nota de Desempenho: Classificador pode ter desempenho diferente em imagens de fêmeas

Conclusão:

- Projetado para: Classificação precisa de espécies de aves
- Precisão Esperada: Alta, especialmente com imagens de machos
- Consideração: Diferenças entre machos e fêmeas podem afetar o desempenho

DINO V2 (Vision Transformer)

DINO V2 é um modelo de Transformer para visão, que se destaca por sua abordagem de treinamento sem supervisão. Ao contrário de métodos tradicionais, utiliza uma abordagem de maximização de similaridade entre representações de diferentes instâncias de uma imagem.



Visualização da Representação

Análise PCA: Para entender a representação do modelo, aplicamos PCA (Principal Component Analysis) para reduzir as dimensões da representação de características.

Segmentação com PCA: Utilizamos os resultados da PCA para separar efetivamente o fundo e o primeiro plano da imagem, permitindo uma análise mais focada nas características relevantes.

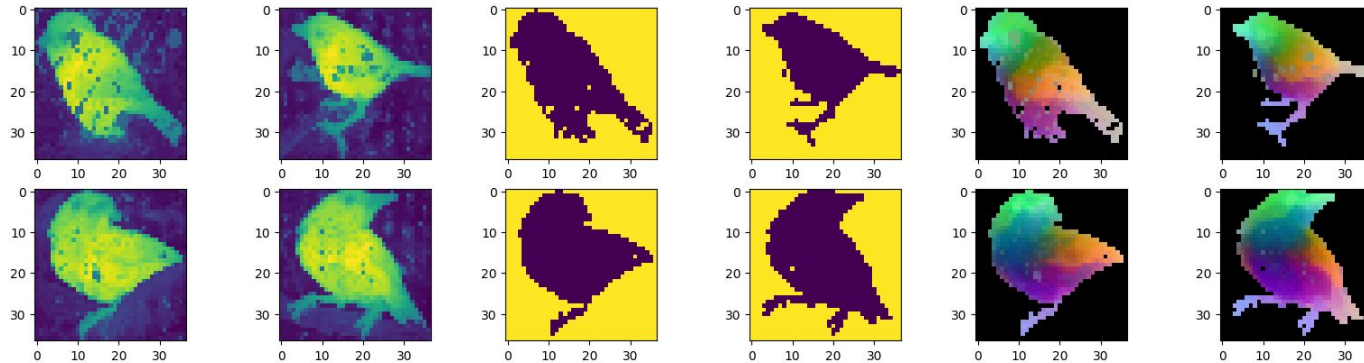
Preparação do Conjunto de Dados:

Dataset Utilizado: Utilizamos um conjunto de dados de imagens classificadas em diferentes categorias, retirado do site kaggle onde apresenta um problema de classificação de espécies de aves.

Transformações de Dados: Aplicamos transformações como redimensionamento, corte central, normalização e aumento de dados para preparar o conjunto de dados para treinamento e teste.

Arquitetura do Modelo:

DinoVisionTransformerClassifier: Implementamos um classificador utilizando a arquitetura DINO V2 como base. O modelo inclui um transformer para extração de características e uma camada densa para classificação.



Treinamento do Modelo:

Otimização: Utilizamos o otimizador Adam com uma taxa de aprendizado reduzida para treinar o modelo.

Épocas e Lotes: O treinamento ocorre ao longo de várias épocas, com atualizações de parâmetros feitas em lotes de dados.

Avaliação de Desempenho: A precisão e a perda são monitoradas durante o treinamento para avaliar o desempenho do modelo.

Teste e Avaliação:

Acurácia: Avaliamos a acurácia do modelo em um conjunto de teste separado.

Matriz de Confusão: Utilizamos uma matriz de confusão para entender como o modelo classifica diferentes categorias.

Relatório de Classificação: Apresentamos um relatório de classificação detalhado para cada classe, incluindo métricas como precisão, recall e F1-score.

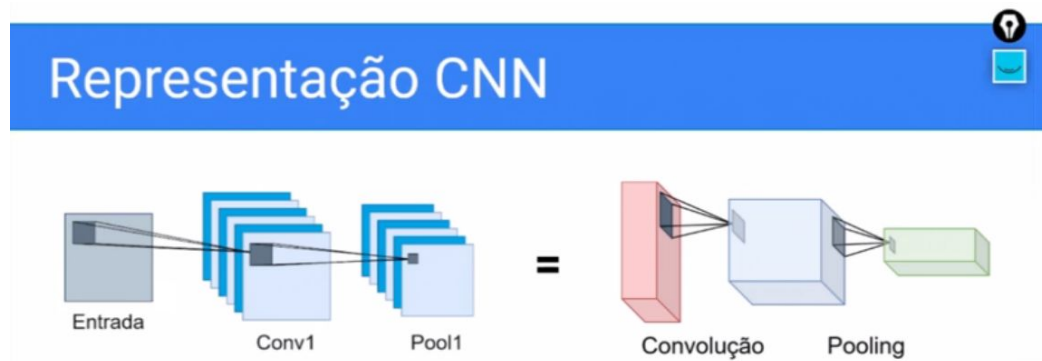
Considerações sobre o DINO V2:

Treinamento Sem Supervisão: O DINO V2 destaca-se por seu treinamento sem supervisão, promovendo uma abordagem inovadora na extração de características em visão computacional.

Representações de Imagens: A aplicação de PCA permitiu uma análise mais profunda das representações de imagens geradas pelo modelo, proporcionando insights sobre a separação de fundo e primeiro plano.

Arquitetura da Rede Neural Convolucional (CNN)

A Convolutional Neural Network (CNN) é um modelo de aprendizado profundo especialmente projetado para processar dados de grade, como imagens. No modelo que será apresentado, utilizamos uma CNN simples para tarefas de classificação de imagens.



Preparação do Conjunto de Dados:

Dataset Utilizado: Utilizamos um conjunto de dados de imagens classificadas em diferentes categorias, retirado do site kaggle onde apresenta um problema de classificação de espécies de aves.

Transformações de Dados: Aplicamos transformações, como redimensionamento, corte central, normalização e aumento de dados, para preparar o conjunto de dados para treinamento e teste.

Arquitetura do Modelo CNN:

Camadas Convolutivas: A CNN possui camadas convolutivas que são capazes de aprender características locais da imagem.

Camadas de Pooling: Utilizamos camadas de pooling para reduzir a dimensionalidade e extrair características mais importantes.

Camadas Totalmente Conectadas: Incluimos camadas totalmente conectadas para a fase final de classificação.

Treinamento do Modelo:

Otimização: Utilizamos o otimizador Adam com uma taxa de aprendizado reduzida para treinar o modelo.

Épocas e Batches: O treinamento ocorre ao longo de várias épocas, com atualizações de parâmetros feitas em lotes de dados.

Avaliação de Desempenho: A precisão e a perda são monitoradas durante o treinamento para avaliar o desempenho do modelo.

Visualização do Treinamento:

Curvas de Loss e Precisão: Apresentamos gráficos mostrando a variação da loss e da precisão durante o treinamento, permitindo uma análise visual do aprendizado do modelo.

Teste e Avaliação:

Acurácia: Avaliamos a acurácia do modelo em um conjunto de teste separado.

Matriz de Confusão: Utilizamos uma matriz de confusão para entender como o modelo classifica diferentes categorias.

Relatório de Classificação: Apresentamos um relatório de classificação detalhado para cada classe, incluindo métricas como precisão, recall e F1-score.

Considerações sobre a CNN:

Aplicações Práticas: Modelos CNN são amplamente utilizados em aplicações de visão computacional, como reconhecimento de objetos, classificação de imagens e detecção de padrões.

Adaptação a Novos Dados: A CNN pode ser adaptada a diferentes conjuntos de dados e tarefas de classificação, tornando-a versátil para diversas aplicações.

obs: grid cv - skit cross validation não foi utilizado.

Treinamentos

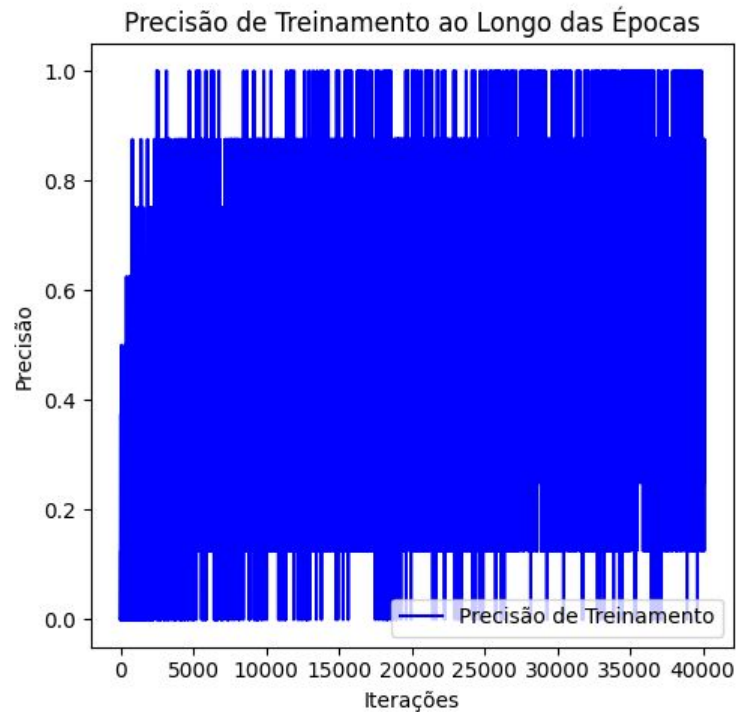
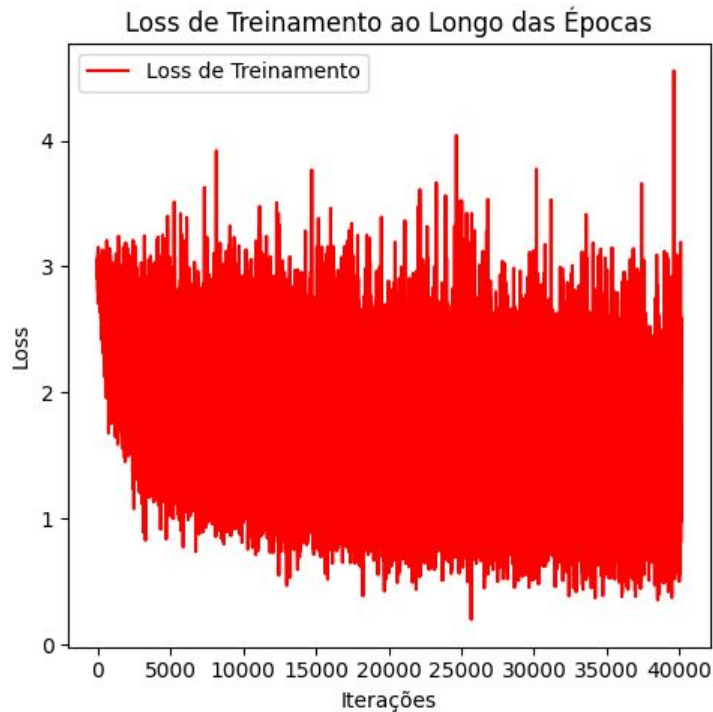
CNN

Epoch [1/100]: 100%		401/401 [00:17<00:00, 23.50it/s, acc=0.25, loss=2.82]
Epoch [2/100]: 100%		401/401 [00:17<00:00, 23.44it/s, acc=0.375, loss=2.1]
Epoch [3/100]: 100%		401/401 [00:16<00:00, 23.99it/s, acc=0.375, loss=2.21]
Epoch [4/100]: 100%		401/401 [00:16<00:00, 24.11it/s, acc=0.25, loss=2.21]
Epoch [5/100]: 100%		401/401 [00:16<00:00, 23.67it/s, acc=0.25, loss=2.56]
Epoch [6/100]: 100%		401/401 [00:17<00:00, 23.56it/s, acc=0.75, loss=1.51]
Epoch [7/100]: 100%		401/401 [00:17<00:00, 23.53it/s, acc=0.5, loss=1.87]
Epoch [8/100]: 100%		401/401 [00:16<00:00, 23.65it/s, acc=0.375, loss=1.98]
Epoch [9/100]: 100%		401/401 [00:17<00:00, 23.02it/s, acc=0.625, loss=1.44]
Epoch [10/100]: 100%		401/401 [00:17<00:00, 22.87it/s, acc=0.375, loss=1.96]
Epoch [11/100]: 100%		401/401 [00:17<00:00, 22.94it/s, acc=0.25, loss=2.21]
Epoch [12/100]: 100%		401/401 [00:17<00:00, 22.88it/s, acc=0.5, loss=1.99]
Epoch [13/100]: 100%		401/401 [00:17<00:00, 22.50it/s, acc=1, loss=0.997]
Epoch [14/100]: 100%		401/401 [00:17<00:00, 22.57it/s, acc=0.5, loss=1.69]
Epoch [15/100]: 100%		401/401 [00:17<00:00, 22.86it/s, acc=0.25, loss=2.24]
Epoch [16/100]: 100%		401/401 [00:17<00:00, 22.89it/s, acc=0.75, loss=1.23]
Epoch [17/100]: 100%		401/401 [00:17<00:00, 22.81it/s, acc=0.875, loss=1.05]
Epoch [18/100]: 100%		401/401 [00:17<00:00, 22.92it/s, acc=0.75, loss=1.53]
Epoch [19/100]: 100%		401/401 [00:17<00:00, 22.69it/s, acc=0.5, loss=1.88]
Epoch [20/100]: 100%		401/401 [00:17<00:00, 22.90it/s, acc=0.75, loss=1.31]
Epoch [21/100]: 100%		401/401 [00:17<00:00, 22.74it/s, acc=0.75, loss=1.74]
Epoch [22/100]: 100%		401/401 [00:17<00:00, 22.73it/s, acc=0.5, loss=1.77]
Epoch [23/100]: 100%		401/401 [00:17<00:00, 22.69it/s, acc=0.625, loss=1.18]
Epoch [24/100]: 100%		401/401 [00:17<00:00, 23.07it/s, acc=0.5, loss=2.03]
Epoch [25/100]: 100%		401/401 [00:17<00:00, 22.82it/s, acc=0.5, loss=2.23]
...		
Epoch [97/100]: 100%		401/401 [00:16<00:00, 23.79it/s, acc=0.5, loss=1.22]
Epoch [98/100]: 100%		401/401 [00:16<00:00, 23.63it/s, acc=0.75, loss=1.28]
Epoch [99/100]: 100%		401/401 [00:16<00:00, 23.75it/s, acc=0.625, loss=1.14]
Epoch [100/100]: 100%		401/401 [00:16<00:00, 24.03it/s, acc=0.25, loss=2.07]

DINO V2

Epoch [0/10]: 100%		401/401 [00:19<00:00, 20.51it/s, acc=0.625, loss=1.9]
Epoch [1/10]: 100%		401/401 [00:18<00:00, 21.61it/s, acc=0.75, loss=1.01]
Epoch [2/10]: 100%		401/401 [00:18<00:00, 21.78it/s, acc=0.625, loss=1.07]
Epoch [3/10]: 100%		401/401 [00:18<00:00, 21.55it/s, acc=1, loss=0.166]
Epoch [4/10]: 100%		401/401 [00:18<00:00, 21.58it/s, acc=1, loss=0.202]
Epoch [5/10]: 100%		401/401 [00:18<00:00, 21.68it/s, acc=0.875, loss=0.413]
Epoch [6/10]: 100%		401/401 [00:18<00:00, 21.65it/s, acc=1, loss=0.0981]
Epoch [7/10]: 100%		401/401 [00:18<00:00, 21.57it/s, acc=1, loss=0.111]
Epoch [8/10]: 100%		401/401 [00:18<00:00, 21.55it/s, acc=0.875, loss=0.4]
Epoch [9/10]: 100%		401/401 [00:18<00:00, 21.52it/s, acc=1, loss=0.0793]

Curvas de Loss e Precisão da CNN



Reports

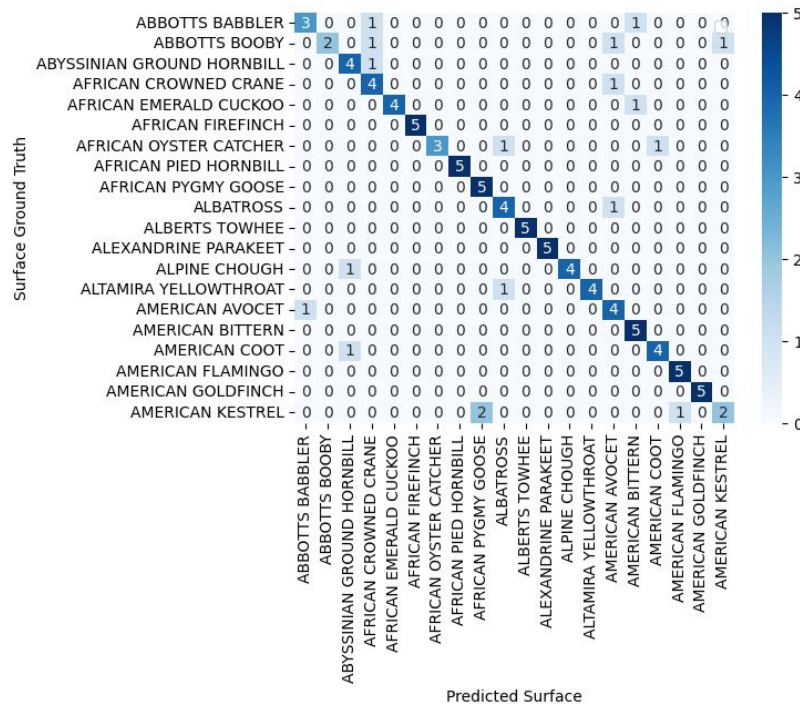
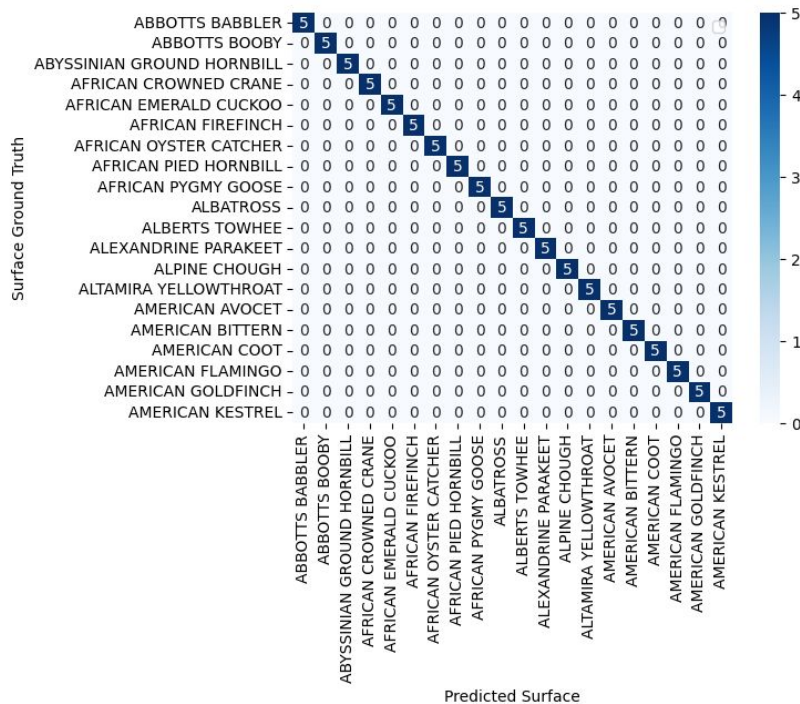
DINO V2

	precision	recall	f1-score	support
ABBOTTS BABBLER	1.00	1.00	1.00	5
ABBOTTS BOOBY	1.00	1.00	1.00	5
ABYSSINIAN GROUND HORNBILL	1.00	1.00	1.00	5
AFRICAN CROWNED CRANE	1.00	1.00	1.00	5
AFRICAN EMERALD CUCKOO	1.00	1.00	1.00	5
AFRICAN FIREFINCH	1.00	1.00	1.00	5
AFRICAN OYSTER CATCHER	1.00	1.00	1.00	5
AFRICAN PIED HORNBILL	1.00	1.00	1.00	5
AFRICAN PYGMY GOOSE	1.00	1.00	1.00	5
ALBATROSS	1.00	1.00	1.00	5
ALBERTS TOWHEE	1.00	1.00	1.00	5
ALEXANDRINE PARAKEET	1.00	1.00	1.00	5
ALPINE CHOUGH	1.00	1.00	1.00	5
ALTAMIRA YELLOWTHROAT	1.00	1.00	1.00	5
AMERICAN AVOCET	1.00	1.00	1.00	5
AMERICAN BITTERN	1.00	1.00	1.00	5
AMERICAN COOT	1.00	1.00	1.00	5
AMERICAN FLAMINGO	1.00	1.00	1.00	5
AMERICAN GOLDFINCH	1.00	1.00	1.00	5
AMERICAN KESTREL	1.00	1.00	1.00	5
accuracy			1.00	100
macro avg	1.00	1.00	1.00	100
weighted avg	1.00	1.00	1.00	100

CNN

Accuracy of the network on the 104 test images: 82 %					
	precision	recall	f1-score	support	
10:36					
ja ?					
ABBOTTS BABBLER	0.75	0.60	0.67	5	
ABBOTTS BOOBY	1.00	0.40	0.57	5	
ABYSSINIAN GROUND HORNBILL	0.67	0.80	0.73	5	
10:37					
AFRICAN CROWNED CRANE	0.57	0.80	0.67	5	
AFRICAN EMERALD CUCKOO	1.00	0.80	0.89	5	
AFRICAN FIREFINCH	1.00	1.00	1.00	5	
AFRICAN OYSTER CATCHER	1.00	0.60	0.75	5	
AFRICAN PIED HORNBILL	1.00	1.00	1.00	5	
AFRICAN PYGMY GOOSE	0.71	1.00	0.83	5	
ALBATROSS	0.67	0.80	0.73	5	
10:38					
ALBERTS TOWHEE	1.00	1.00	1.00	5	
ALEXANDRINE PARAKEET	1.00	1.00	1.00	5	
ALPINE CHOUGH	1.00	0.80	0.89	5	
ALTAMIRA YELLOWTHROAT	1.00	0.80	0.89	5	
10:39					
AMERICAN AVOCET	0.57	0.80	0.67	5	
AMERICAN BITTERN	0.71	1.00	0.83	5	
AMERICAN COOT	0.80	0.80	0.80	5	
AMERICAN FLAMINGO	0.83	1.00	0.91	5	
10:40					
AMERICAN GOLDFINCH	1.00	1.00	1.00	5	
AMERICAN KESTREL	0.67	0.40	0.50	5	
10:41					
accuracy			0.82	100	
10:42					
macro avg	0.85	0.82	0.82	100	
10:43					
weighted avg	0.85	0.82	0.82	100	

DINO V2 vs CNN



Obrigado pela atenção!