

Image Classification on the Modified MPEG7 Dataset Using k-NN and RNN Algorithms*

1st Bernardo Teixeira de Miranda

Dept. de Sistemas de Informação

Universidade de Viçosa

Rio Parnaíba, Brazil

bernardo.t.miranda@ufv.br

2nd Phelipe Romano

Dept. de Sistemas de Informação

Universidade de Viçosa

Rio Parnaíba, Brazil

phelipe.romano@ufv.br

3rd Vinícius do Carmo Gomes

Dept. de Sistemas de Informação

Universidade de Viçosa

Rio Parnaíba, Brazil

vinicius.c.gomes@ufv.br

Abstract—This project focuses on the classification of images using the modified MPEG7 dataset. The methodology encompasses several stages, including image segmentation, morphological feature extraction, data normalization, and classifier performance analysis. Initially, the k-Nearest Neighbors (k-NN) algorithm was employed, chosen for its simplicity and effectiveness in handling classification tasks based on proximity. The k-NN served as a baseline model, demonstrating robust performance with fundamental morphological features. Subsequently, a Recurrent Neural Network (RNN) was implemented to explore its ability to capture complex patterns and sequential relationships within the dataset. However, despite its potential to learn advanced representations and address non-linear dependencies, the RNN showed inferior performance compared to the k-NN under the given conditions.

Key steps in the project included dataset partitioning into training, validation, and testing sets, along with preprocessing such as data normalization. The models were evaluated using metrics like accuracy, precision, recall, and F1-score. The results highlighted the comparative advantages of using RNN over k-NN, underscoring the importance of selecting appropriate algorithms for complex classification tasks. The project concludes with a comprehensive analysis of the results and provides documentation to guide reproducibility through a GitHub repository.

Index Terms—component, formatting, style, styling, insert

I. INTRODUCTION

A classificação de imagens é uma área essencial na ciência de dados, com aplicações em diversos domínios, como diagnóstico médico, sistemas de vigilância e automação. Este projeto tem como objetivo principal desenvolver e avaliar classificadores para o conjunto de dados MPEG7 modificado, aplicando uma sequência estruturada de etapas que abrangem desde a segmentação das imagens até a análise de desempenho dos modelos.

Inicialmente, foi utilizado o algoritmo k-NN (K-Nearest Neighbors), sendo destacado pela sua simplicidade e eficácia em problemas de classificação baseados na proximidade dos dados. O k-NN é um modelo não paramétrico que utiliza a distância entre amostras para atribuir classes, sendo uma boa escolha para estabelecer um ponto de partida no projeto e avaliar o desempenho inicial com características morfológicas extraídas das imagens.

Posteriormente, foi implementada uma Rede Neural Recorrente (RNN), capaz de capturar relações sequenciais e padrões complexos nos dados. Embora a RNN tenha o potencial

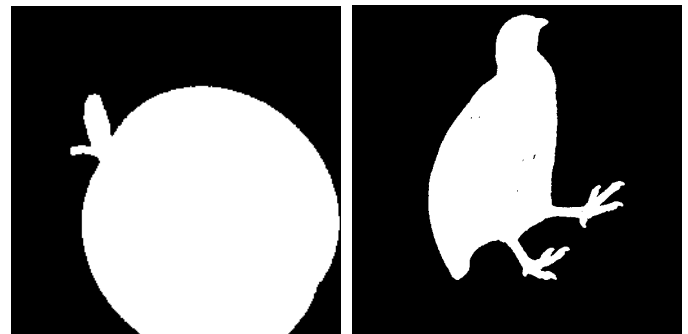
de aprender representações mais avançadas, indo além das limitações do k-NN em cenários com dependências não lineares ou características que exigem maior capacidade de generalização, ela apresentou um desempenho inferior ao k-NN.

O relatório detalha cada etapa do projeto, desde a extração de características morfológicas até a divisão dos dados em conjuntos de treino e teste, além da normalização dos dados. Os resultados obtidos são analisados com base em métricas como acurácia, precisão, recall e F1-score, demonstrando a evolução de desempenho ao longo do uso dos diferentes classificadores.

II. METODOLOGIA

A. Conjuntos de Imagens

- 1) **MPEG7 Modificado:** O dataset MPEG7 modificado é composto por imagens de objetos e cenas organizadas em várias classes. Ele contém imagens com diferentes resoluções e condições, adequadas para tarefas de segmentação e classificação. As imagens têm uma resolução variável e estão armazenadas em um formato padrão. A imagem 1 mostra um exemplo de duas classes morfológicas presentes no dataset.



(a) Imagem de uma maçã

(b) Imagem de um pássaro

Fig. 1: Exemplo de classes do dataset MPEG7 modificado: (a) apple, (b) bird

B. Coleta e Preparação de Dados

Esta etapa envolve a coleta e preparação dos dados da base MPEG7, modificados para serem utilizados no treinamento e validação dos modelos. Inicialmente, as imagens foram coletadas de um diretório contendo subpastas organizadas em seis classes distintas: Apple, Bat, Beetle, Bell, Bird e Bone. Cada classe continha 100 figuras binárias com diferentes tamanhos, posições e resoluções. Essas imagens foram rotuladas com base na subpasta de origem e redimensionadas para 1/4 do tamanho original, resultando em uma redução da complexidade operacional e tempo gasto para treino.

C. Pré-Processamento

O pré-processamento envolveu a extração de características para a identificação de padrões e classificação do modelo. Primeiramente foi realizada a segmentação das imagens, isolando o maior objeto encontrado e preenchendo eventuais buracos. Em seguida, foi realizada a identificação das características morfológicas: Área, eixo maior, eixo menor, solidez e excentricidade. Estes dados foram organizados em uma tabela e posteriormente normalizados. Essas medidas facilitaram o processo de aprendizado e classificação do modelo, melhorando a qualidade dos dados e destacando informações mais relevantes.

D. Treinamento e Avaliação das Redes Neurais Convolucionais

Inicialmente, o conjunto de dados foi dividido de forma estratificada, destinando 70% para treinamento e 30% para teste, garantindo a representatividade proporcional das classes em ambos os subconjuntos.

As características morfológicas selecionadas para a classificação foram a área e o maior eixo, devido à sua maior relevância na identificação de padrões e na discriminação entre as classes. Para a configuração dos modelos, o k-NN foi definido para classificar as amostras com base nos três vizinhos mais próximos, enquanto o RNN utilizou um raio de proximidade igual a 1 para determinar os grupos de referência.

Os parâmetros escolhidos para ambas as técnicas foram ajustados para maximizar o desempenho no conjunto de dados utilizado, equilibrando simplicidade e eficácia.

E. Métricas

Para a análise dos resultados foram utilizadas formulas para encontrar a precisão (Prec), recall (Recall), F1-score (F1) e support (Sup) para cada classe dos testes realizados com base nos resultados da matriz de confusão. As formulas também foram utilizadas para avaliar o desempenho dos modelos por meio da media aritimetica do resultado das classes. Além disso, foi calculado o valor da acurácia (Acc) dos modelos.

III. CONCLUSÃO

A partir dos resultados obtidos pelos modelos K-NN e R-NN, representados pelas imagens 2 e 3, respectivamente, pode-se observar que, em todas as classes morfológicas, o modelo k-NN apresentou desempenho superior, alcançando

uma acurácia de 82%. Em contrapartida, o modelo RNN obteve uma acurácia de 61%.

Essa diferença significativa de 21% na acurácia indica que, para as classes morfológicas utilizadas neste estudo, o modelo k-NN se mostrou mais eficiente com os parâmetros aplicados. Portanto, este artigo verificou que o k-NN é uma escolha mais adequada para a classificação com base nas características morfológicas exploradas quando comparada ao modelo RNN.

Matriz de confusão:

[[29	0	0	1	0	0]	
[0	22	0	0	8	0]
[0	0	23	0	5	2]
[4	0	0	26	0	0]
[0	7	2	0	21	0]
[0	0	4	0	1	25]

Relatório de classificação:

	precision	recall	f1-score	support
apple	0.88	0.97	0.92	30
bat	0.76	0.73	0.75	30
beetle	0.79	0.77	0.78	30
bell	0.96	0.87	0.91	30
bird	0.60	0.70	0.65	30
bone	0.93	0.83	0.88	30
accuracy			0.81	180
macro avg	0.82	0.81	0.81	180
weighted avg	0.82	0.81	0.81	180

Fig. 2: Matriz de confusão e relatório de classificação do modelo k-NN

Matriz de confusão:

```
[[30  0  0  0  0  0]
 [ 2 22  3  0  3  0]
 [ 1  2  9  1  5 12]
[17  0  0 13  0  0]
 [ 1  8 11  1  9  0]
 [ 0  0  3  0  1 26]]
```

Relatório de classificação:

	precision	recall	f1-score	support
apple	0.59	1.00	0.74	30
bat	0.69	0.73	0.71	30
beetle	0.35	0.30	0.32	30
bell	0.87	0.43	0.58	30
bird	0.50	0.30	0.38	30
bone	0.68	0.87	0.76	30
accuracy			0.61	180
macro avg	0.61	0.61	0.58	180
weighted avg	0.61	0.61	0.58	180

Fig. 3: Matriz de confusão e relatório de classificação do modelo RNN

REFERENCES

- [1] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Pearson, 2007.
- [2] L. da F. Costa and R. M. Cesar-Jr., *Shape Analysis and Classification: Theory and Practice*, CRC Press, 2000, Chapter 8.
- [3] Scikit-image Documentation, Available at: <https://scikit-image.org/docs/stable/>.
- [4] Scikit-learn, *User Guide*, Available at: https://scikit-learn.org/stable/user_guide.html.