

# Extração de características de imagens usando HOG e CNN e treinamento do modelo usando $k$ -NN

Luiz Felipe de Souza Silva<sup>1</sup>

<sup>1</sup>Departamento de Engenharia da Computação (DCA) – Universidade Federal do Rio Grande do Norte (UFRN)

1f06092004@gmail.com

**Resumo.** Este relatório apresenta uma análise comparativa dos métodos de extração de características HOG e CNN aplicados à tarefa de classificação de imagens de raças de cães e gatos (Egyptian Mau, Ragdoll, Pomeranian e Shiba Inu), utilizando o algoritmo  $k$ -NN como classificador. Também são discutidos os impactos da redução de dimensionalidade por PCA e da variação do parâmetro  $k$  no desempenho do modelo.

## 1. Informações Gerais.

O objetivo da atividade foi realizar a classificação de raças de gatos e cachorros a partir de imagens, utilizando o classificador  $k$ -NN. As características das imagens foram extraídas por meio dos métodos HOG e CNNs pré-treinadas (VGG16 e VGG19). Após o treinamento, os resultados foram utilizados para responder a três questões-chave relacionadas ao desempenho dos métodos.

1. Qual foi o melhor método de extração HOG ou CNN? Tente explicar o motivo do comportamento da sua base de dados
2. O que aconteceu com a acurácia quando diminuímos o número de atributos com o PCA? Explique o comportamento dos modelos.
3. Em relação ao  $k$ -NN, o que aconteceu quando aumentamos o número de vizinhos? Explique o motivo deste comportamento

Os dados de acurácia do modelo treinado com  $k$ -NN podem ser visto através desse direcionamento para tabela: [Tabela\\_k-NN\\_Checkpoint 1](#)

Notebooks, extração de características e documentação podem ser visto no repositório do github: [Pré Processamento e k-NN](#)

### 1.1 HOG ou CNN

De acordo com os resultados apresentados, as redes CNNs pré-treinadas (VGG16 e VGG19) se mostraram os melhores extratores de características quando comparadas ao método HOG. Isso se deve ao fato que as CNNs possuem mais camadas convolucionais e são capazes de aprender de forma mais discriminativa as características das classes favorecendo os classificadores como o  $k$ -NN.

A melhor extração de HOG não apresenta metade da média da pior CNN. A melhor extração com HOG foi obtida com imagens padronizadas em 128x128 pixels e blocos de 20x20, alcançando uma média de treinamento, enquanto a pior extração com CNN utilizando pooling “avg” ou “max” com padronização de imagem de 128x128 pixels

apresentou uma acurácia de 93,14%. Ou seja, a melhor extração de HOG não equivale a metade da pior extração de CNN, evidenciando a superioridade das redes convolucionais nesse aspecto.

## 1.2 PCA favorável?

A aplicação do PCA foi realizada nas seis melhores bases: CNN\_VGG16\_256\_max, CNN\_VGG16\_256\_avg, CNN\_VGG19\_128\_max, CNN\_VGG19\_128\_avg, CNN\_VGG19\_256\_max e CNN\_VGG19\_256\_avg.

Com a aplicação do PCA, podem-se esperar dois comportamentos: a acurácia pode aumentar, caso o PCA consiga eliminar redundâncias e selecionar as características mais discriminantes das classes; ou pode diminuir, caso o PCA elimine muitos componentes importantes, necessários para uma boa classificação.

Nas bases especificadas anteriormente, o uso PCA foi favorável, aumentando a acurácia em alguns pontos percentuais, eliminando redundâncias e conseguindo discriminar muito bem as classes, com alta eficiência. A base que obteve o maior ganho percentual foi à base que utilizou a CNN pré treinada VGG 16 com imagens padronizadas em 256x256 pixels com entradas de pooling “max”, 1,09%.

**Tabela 1:** Comparação após a aplicação do PCA.

Base	Acurácia Média (Sem PCA)	Acurácia Média (Com PCA)
CNN_VGG16_256_max	98,05%	99,14%
CNN_VGG16_256_avg	98,37%	98,80%
CNN_VGG19_128_max	93,17%	93,85%
CNN_VGG19_128_avg	93,82%	94,29%
CNN_VGG19_256_max	98,49%	98,91%
CNN_VGG19_256_avg	97,22%	98,26%

## 1.3 Comportamento do $k$

Com relação ao comportamento do número de vizinhos, é perceptível notar um padrão não linear, com oscilações conforme a quantidade de vizinhos utilizada. Nesta análise, optou-se por desconsiderar o valor de  $k = 1$ , por ser considerado muito baixo para uma predição adequada, podendo levar a resultados inconsistentes. Analisando a média geral das acurácias em função da quantidade de vizinhos, observa-se que o intervalo ideal está

entre  $k = 4$  e  $k = 8$ . É neste intervalo que a média geral das acurácias apresenta um ponto de máximo, tanto antes quanto após a aplicação do PCA.

Ao se considerar valores de  $k$  acima desse intervalo, nota-se um decaimento gradual da acurácia, o que pode ser um indicativo de que uma quantidade elevada de vizinhos causa confusão no modelo. Isso ocorre porque, ao considerar muitos vizinhos, o algoritmo tende a incluir amostras de classes diferentes, prejudicando a assertividade da predição. Por outro lado, ao se utilizar valores de  $k$  abaixo desse intervalo, também se observa uma redução das acurácias, sugerindo que uma quantidade muito pequena de vizinhos não é suficiente para o modelo realizar uma boa generalização. Com o gráfico a seguir é notável ver essa estabilidade no intervalo indicado, evidenciado ainda mais que para esse problema, o  $k$  entre 4 e 8 é o ideal para melhor acurácia do modelo.

