Sistema de Busca e Classificação de Animais de Estimação

Caue Lira

3 de junho de 2026

Resumo

Este trabalho apresenta um sistema de busca e classificação de animais de estimação baseado em descritores clássicos de imagens. Foram avaliados seis descritores globais ou semiglobais, incluindo um descritor não visto em aula, além de combinações entre eles. A base utilizada contém 367 imagens distribuídas em 43 classes de pets. Para classificação, foram consideradas 362 imagens de 38 classes, removendo pets com menos de três fotos. Para busca, foram usadas todas as 367 imagens. Os experimentos mostram que a combinação de todos os descritores obteve o melhor resultado de classificação, enquanto uma combinação compacta de cor e textura, GCH+LBP+GLCM, produziu o melhor resultado de busca. O método Bag of Visual Words também foi implementado e analisado por métricas de busca e projeções em duas dimensões.

1 Introdução

A identificação visual de animais de estimação é uma tarefa difícil para descritores clássicos, pois diferentes imagens do mesmo pet podem variar em pose, iluminação, enquadramento, fundo e escala. Ao mesmo tempo, pets diferentes podem compartilhar cores de pelagem, texturas e formas muito parecidas. Neste trabalho, o objetivo é investigar até que ponto descritores de cor, textura e forma conseguem representar características úteis para duas tarefas: classificação do pet e busca por imagens similares.

O estudo foi organizado em torno de três perguntas experimentais. Primeiro, quais tipos de descritores capturam melhor as diferenças entre pets individuais? Segundo, combinações de descritores são mais robustas do que descritores isolados? Terceiro, uma representação por Bag of Visual Words (BoVW) consegue competir com os descritores globais usados diretamente?

Para responder a essas perguntas, foram extraídos seis descritores: histograma global de cores em HSV (GCH), Local Binary Patterns (LBP), atributos de Haralick derivados de matrizes de co-ocorrência (GLCM), Histogram of Oriented Gradients (HOG), correlograma de cores e filtros de Gabor. O descritor de Gabor foi escolhido como o descritor não visto em aula. Os mesmos descritores foram usados em classificação e busca, permitindo comparar como cada representação se comporta em tarefas semânticas diferentes.

2 Base de Dados e Protocolo Experimental

A base recebida contém um arquivo CSV com metadados e duas versões das imagens. O arquivo `pets.csv` lista o identificador da classe, o nome padronizado do pet e o caminho relativo da imagem. A pasta `pets256` contém as imagens redimensionadas para 256x256 pixels e com nomes padronizados; essa foi a versão usada nos experimentos. A pasta `pets_original` mantém os nomes, dimensões e formatos originais, servindo apenas como referência.

Embora o enunciado mencione 42 animais de estimação, o CSV disponibilizado contém 367 imagens distribuídas em 43 classes. Como o relatório deve refletir os dados efetivamente usados, todos os números apresentados a seguir são calculados a partir do CSV e dos arquivos fornecidos.

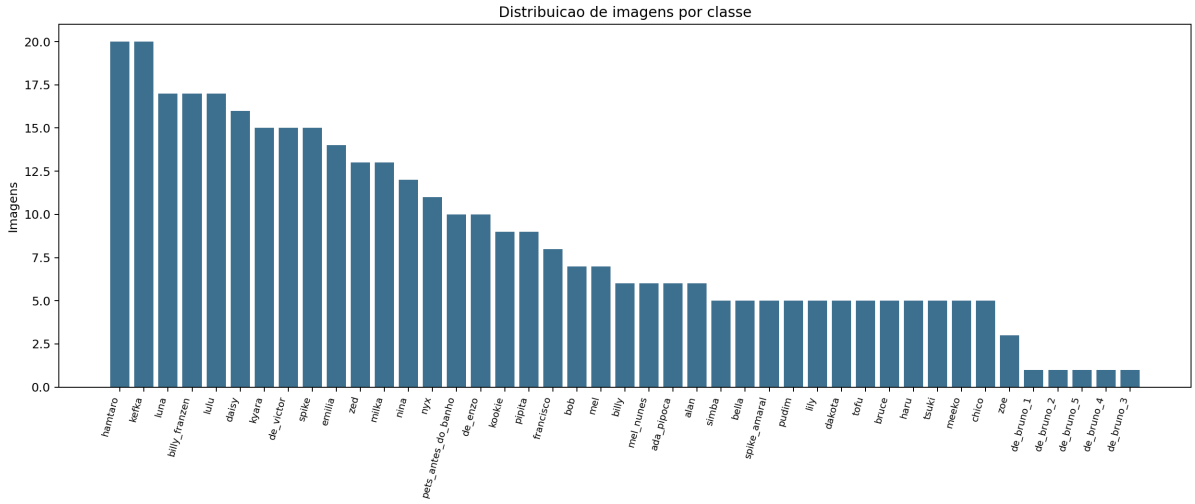


Figura 1: Distribuição do número de imagens por classe na base de dados.

Para a tarefa de classificação, pets com menos de três fotos foram excluídos, como definido no enunciado. Isso removeu cinco classes com uma imagem cada: `de_bruno_1`, `de_bruno_2`, `de_bruno_3`, `de_bruno_4` e `de_bruno_5`. Assim, a classificação usou 362 imagens de 38 classes. A divisão seguiu fielmente a proporção global solicitada de 80%, 10% e 10%, resultando em 290 imagens de treino, 36 de validação e 36 de teste. A separação entre treino e subconjunto temporário foi estratificada por classe; em seguida, o subconjunto temporário foi dividido igualmente em validação e teste. Todos os sorteios usaram semente fixa igual a 42, para tornar os experimentos reprodutíveis. Como validação e teste têm apenas 36 imagens cada em um problema com 38 classes, as métricas desses conjuntos devem ser interpretadas como estimativas de alta variância.

O classificador escolhido foi uma SVM, com seleção de hiperparâmetros no conjunto de validação. Foram avaliados kernel linear com valores de $C \in \{0.1, 1, 10\}$ e kernel RBF com $C \in \{0.1, 1, 10\}$ e $\gamma \in \{\text{scale}, 0.01, 0.001\}$. A configuração com maior acurácia de validação foi usada para avaliar o conjunto de teste.

Na tarefa de busca, todas as 367 imagens foram usadas. Para cada imagem de consulta, o sistema calcula a distância euclidiana entre o vetor da consulta e os vetores das demais imagens, removendo a própria consulta do ranking. Os resultados foram avaliados por Precision@1, Precision@5, Precision@10 e mean Average Precision (mAP). Como a busca é uma tarefa não supervisionada sobre a própria coleção, não foi aplicado split de treino, validação e teste nessa etapa.

Antes de treinar a SVM ou calcular distâncias euclidianas, os vetores foram padronizados. Em combinações de descritores, cada bloco foi normalizado separadamente antes da concatenação. Essa decisão evita que descritores de maior escala numérica ou maior dimensionalidade dominem o classificador e a distância.

3 Descritores: Semântica, Implementação e Hipóteses

3.1 GCH: Histograma Global de Cores

O histograma global de cores busca representar a distribuição cromática da imagem. A semântica esperada é simples: pets com pelagens muito distintas, ou fotografados em fundos visualmente diferentes, devem produzir histogramas diferentes. Esse descritor foi escolhido porque a cor costuma ser uma das primeiras pistas visuais usadas para distinguir animais, principalmente quando há diferenças claras entre pelagens claras, escuras, alaranjadas, brancas ou manchadas.

Na implementação, cada imagem RGB foi convertida para HSV. O canal de matiz foi quantizado em 16 intervalos, saturação em 4 e valor em 4, totalizando $16 \times 4 \times 4 = 256$ bins. O histograma final foi normalizado por soma L1.

A hipótese para classificação era que o GCH teria desempenho razoável, mas limitado, pois ele ignora a posição das cores e não diferencia bem pets de pelagem semelhante. Para busca, a hipótese era semelhante: o descritor deveria retornar bons vizinhos quando a cor dominante fosse discriminante, mas poderia falhar quando o fundo ou a iluminação dominassem a distribuição de cores.

3.2 LBP: Local Binary Patterns

O LBP descreve padrões locais de textura ao comparar cada pixel com seus vizinhos. A semântica pretendida é capturar regularidades finas da imagem, como textura da pelagem, presença de pelos longos ou curtos, rugosidade e pequenos padrões locais. Esse descritor foi escolhido porque textura é uma característica complementar à cor: dois pets podem ter cores parecidas, mas padrões de pelo diferentes.

A implementação usou a imagem em escala de cinza e uma vizinhança 3x3. Para cada pixel interno, os oito vizinhos foram comparados com o valor central, formando um código binário de 8 bits. O vetor final é um histograma normalizado com 256 bins.

A hipótese para classificação era que o LBP poderia ser útil, mas não necessariamente forte sozinho, por ser sensível a iluminação, pose e escala. Para busca, esperava-se que o LBP fosse mais competitivo quando combinado com cor, pois a textura local poderia refinar resultados visualmente semelhantes em termos cromáticos.

3.3 GLCM/Haralick

As matrizes de co-ocorrência de níveis de cinza modelam relações estatísticas entre intensidades próximas. A semântica capturada é textura global: contraste, regularidade, homogeneidade, energia e complexidade da distribuição de tons. Esse descritor foi escolhido como uma alternativa estatística ao LBP, menos focada em microcódigos locais e mais voltada a propriedades agregadas.

Na implementação, a imagem em cinza foi quantizada em 32 níveis. Foram calculadas matrizes de co-ocorrência para distâncias 1, 2 e 3 e quatro direções. Para cada

matriz, extraíram-se contraste, homogeneidade, energia, correlação e entropia. O vetor final contém a média e o desvio padrão dessas propriedades, resultando em 10 dimensões.

A hipótese para classificação era que o GLCM poderia funcionar como descritor compacto, mas dificilmente capturaria detalhes suficientes para separar muitos pets individuais. Para busca, esperava-se desempenho intermediário, com maior utilidade em combinação com descritores de cor.

3.4 HOG: Histogram of Oriented Gradients

O HOG representa a distribuição de orientações de gradientes, sendo associado a forma, contornos e estrutura de bordas. Sua semântica esperada está relacionada à silhueta do animal, pose, orelhas, cabeça, patas e outras estruturas geométricas. Ele foi incluído porque forma é uma dimensão visual distinta de cor e textura.

As imagens foram convertidas para cinza e redimensionadas para 128x128 pixels. Em seguida foram calculados gradientes por filtros de Sobel, organizados em células de 16x16 pixels com 9 bins de orientação. Blocos 2x2 foram normalizados por norma L2, gerando um vetor de 1764 dimensões.

A hipótese para classificação era ambígua: HOG poderia ajudar quando a pose fosse consistente, mas também poderia sofrer com variações grandes de enquadramento e posição do pet. Para busca, esperava-se que a alta dimensionalidade e a variação de pose dificultassem a distância euclidiana, tornando o descritor frágil isoladamente.

3.5 Correlograma de Cores

O correlograma de cores adiciona informação espacial à cor. Em vez de apenas contar cores, ele mede a frequência com que pixels de mesma cor aparecem a certas distâncias. A semântica pretendida é capturar padrões como manchas, distribuição espacial da pelagem e repetição de cores em regiões próximas.

Na implementação, a imagem foi convertida para HSV e o canal de matiz foi quantizado em 16 níveis. Para as distâncias 1, 3 e 5, foram calculadas frequências de repetição da mesma cor em deslocamentos horizontais, verticais e diagonais. O vetor final tem $16 \times 3 = 48$ dimensões.

A hipótese para classificação era que o correlograma poderia melhorar em relação ao histograma global quando os padrões espaciais de cor fossem importantes. Por outro lado, havia o risco de perder informação por usar apenas matiz e de sofrer com mudanças de pose. Para busca, esperava-se que fosse útil em imagens com manchas e distribuição de cores característica.

3.6 Filtros de Gabor

Os filtros de Gabor foram escolhidos como o descritor não visto em aula. Eles modelam respostas a padrões de textura em diferentes frequências e orientações, capturando estruturas como listras, pelos orientados, transições repetidas e padrões multi-escala. Sua semântica é especialmente ligada a textura orientada.

A implementação converte a imagem para cinza e a redimensiona para 128x128 pixels. Foram usados filtros em 5 frequências e 8 orientações. Para cada resposta, foram extraídos média, desvio padrão e energia, produzindo um vetor de 120 dimensões.

A hipótese para classificação era que Gabor poderia ser um descritor individual forte, porque combina textura, orientação e escala em uma representação relativamente compacta. Para busca, esperava-se que funcionasse bem em pets com padrões de pelagem consistentes, mas que fosse prejudicado quando o fundo tivesse textura dominante.

4 Resultados Quantitativos

4.1 Classificação

A Tabela 1 apresenta os resultados de classificação. A melhor acurácia de teste foi obtida pela combinação de todos os descritores, com 38,89%. Entre descritores isolados, o HOG obteve a maior acurácia de teste, com 27,78%, seguido por Gabor e Correlograma. Embora esses valores sejam modestos em termos absolutos, eles estão acima de uma referência aleatória uniforme para 38 classes, que teria acurácia esperada de aproximadamente 2,63%. Isso indica que os descritores capturam informação discriminante, ainda que insuficiente para resolver completamente a identificação individual dos pets.

Tabela 1: Resultados de classificação por descritor e combinação.

Método	Modelo	Dim.	Val. Acc	Test Acc	F1 Macro	F1 Pond.
GCH	RBF, C=1	256	0.3333	0.1111	0.0431	0.0671
LBP	RBF, C=10	256	0.5556	0.1667	0.0940	0.1213
GLCM	RBF, C=1	10	0.2500	0.1111	0.0766	0.1111
HOG	Linear, C=0.1	1764	0.4167	0.2778	0.1963	0.2389
Correlograma	RBF, C=1	48	0.2500	0.1944	0.1280	0.1843
Gabor	RBF, C=10, $\gamma = 0.01$	120	0.3056	0.2222	0.1599	0.2378
GCH+LBP	Linear, C=0.1	512	0.5000	0.2778	0.1782	0.2176
GCH+GLCM	RBF, C=10, $\gamma = 0.001$	266	0.3889	0.2222	0.1453	0.1981
GCH+LBP+GLCM	Linear, C=0.1	522	0.5000	0.2778	0.1678	0.1963
GCH+LBP+GLCM+Gabor	Linear, C=0.1	642	0.4722	0.3056	0.2261	0.2447
Todos	Linear, C=0.1	2454	0.5000	0.3889	0.2610	0.3479

A Tabela 2 resume as matrizes de confusão dos seis descritores individuais. Ela apresenta o número de acertos na diagonal e as confusões mais frequentes observadas no teste. As matrizes completas de todos os métodos avaliados, incluindo descritores individuais e combinações, podem ser consultadas na seção final de matrizes de confusão.

Tabela 2: Resumo das matrizes de confusão dos descritores individuais.

Descritor	Test Acc	F1 Macro	Acertos	Principais confusões
GCH	0.1111	0.0431	4/36	`hamtaro`→`de_victor` (2); `ada_pipoca`→`kefka` (1)
LBP	0.1667	0.0940	6/36	`ada_pipoca`→`hamtaro` (1); `alan`→`spike` (1)
GLCM	0.1111	0.0766	4/36	`milka`→`hamtaro` (2); `ada_pipoca`→`luna` (1)
HOG	0.2778	0.1963	10/36	`hamtaro`→`emilia` (3); `ada_pipoca`→`lulu` (1)
Correlograma	0.1944	0.1280	7/36	`bob`→`de_victor` (2); `lulu`→`hamtaro` (2)
Gabor	0.2222	0.1599	8/36	`ada_pipoca`→`hamtaro` (1); `alan`→`kefka` (1)

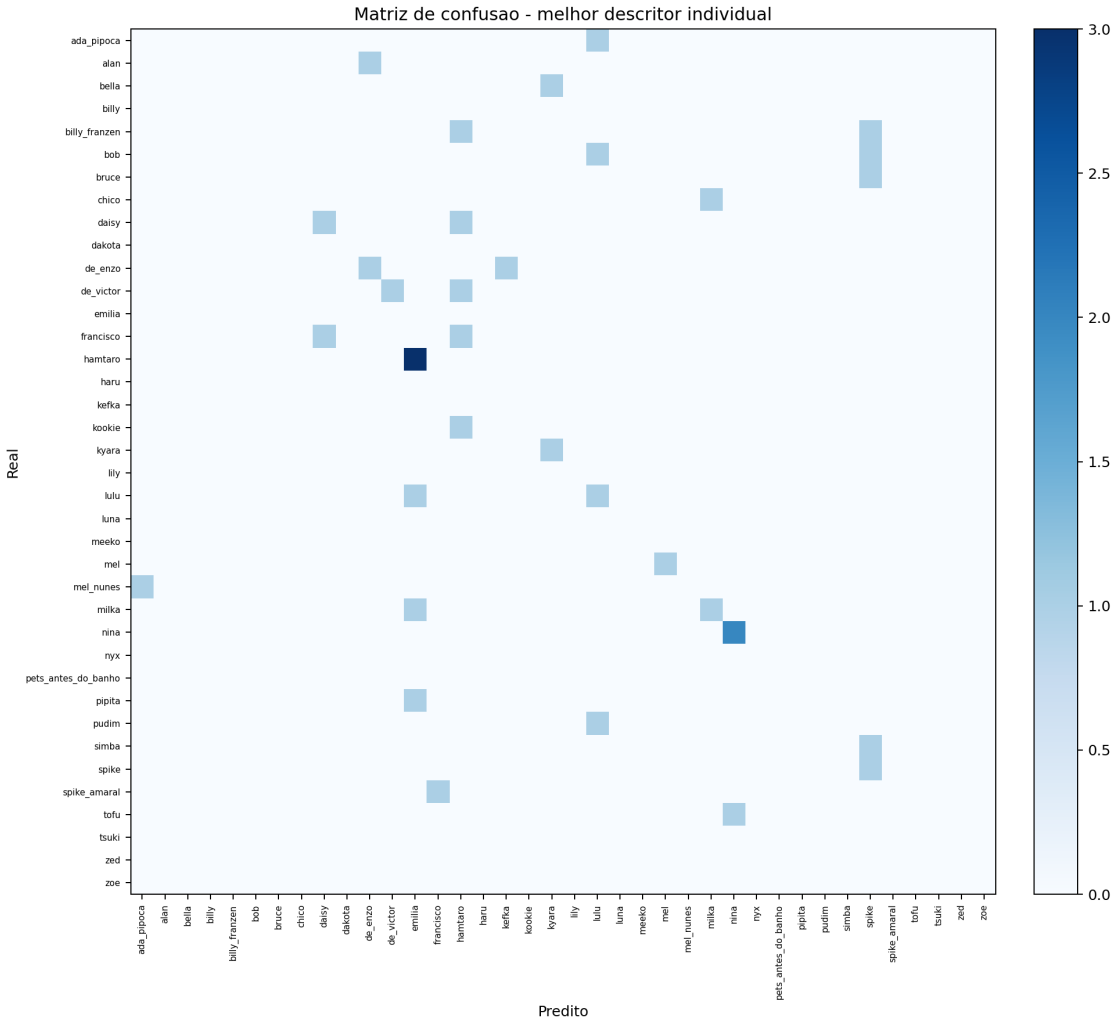


Figura 2: Matriz de confusão para HOG, melhor descritor individual na classificação (Test Acc = 27,78%).

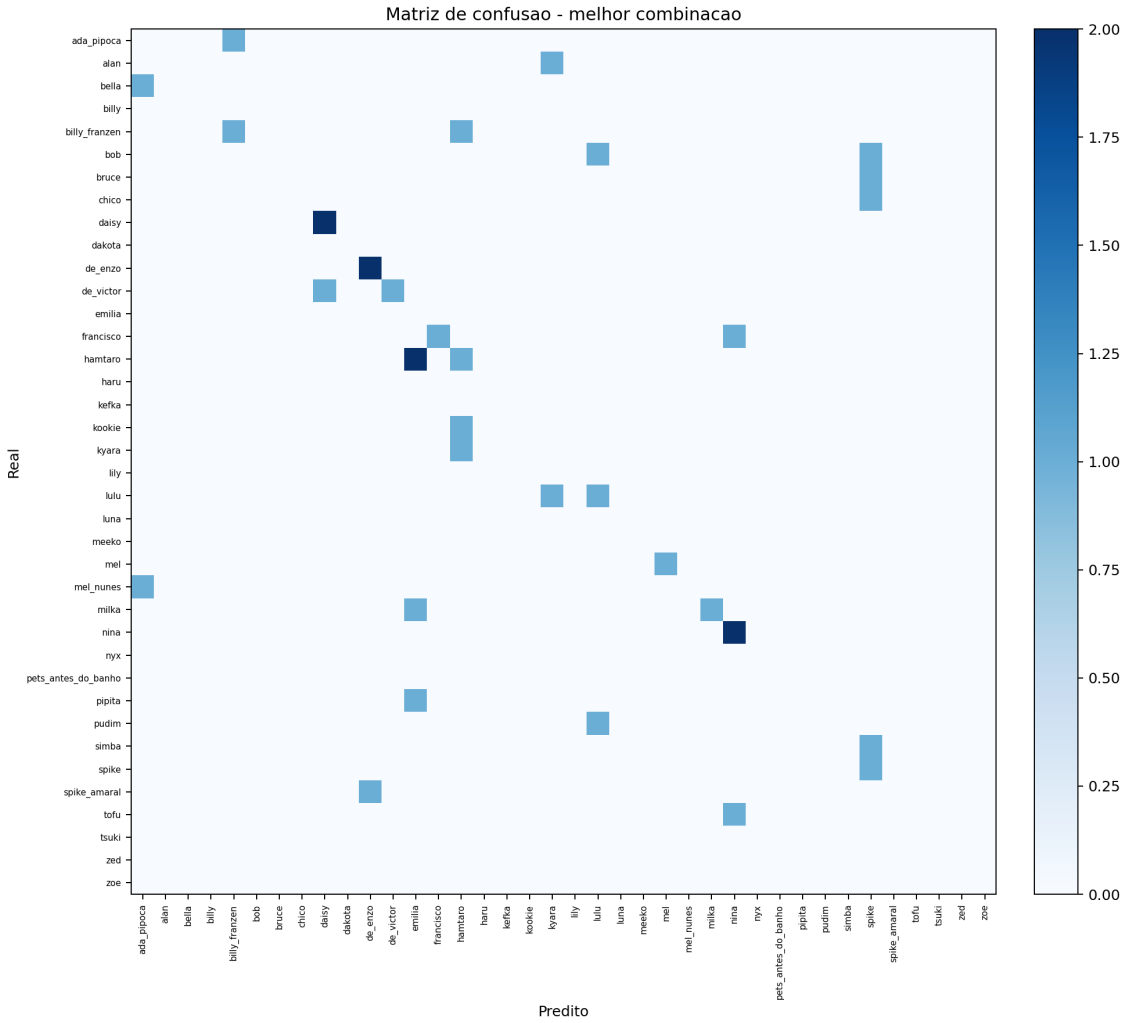


Figura 3: Matriz de confusão para a combinação de todos os descritores, melhor resultado de classificação (Test Acc = 38,89%).

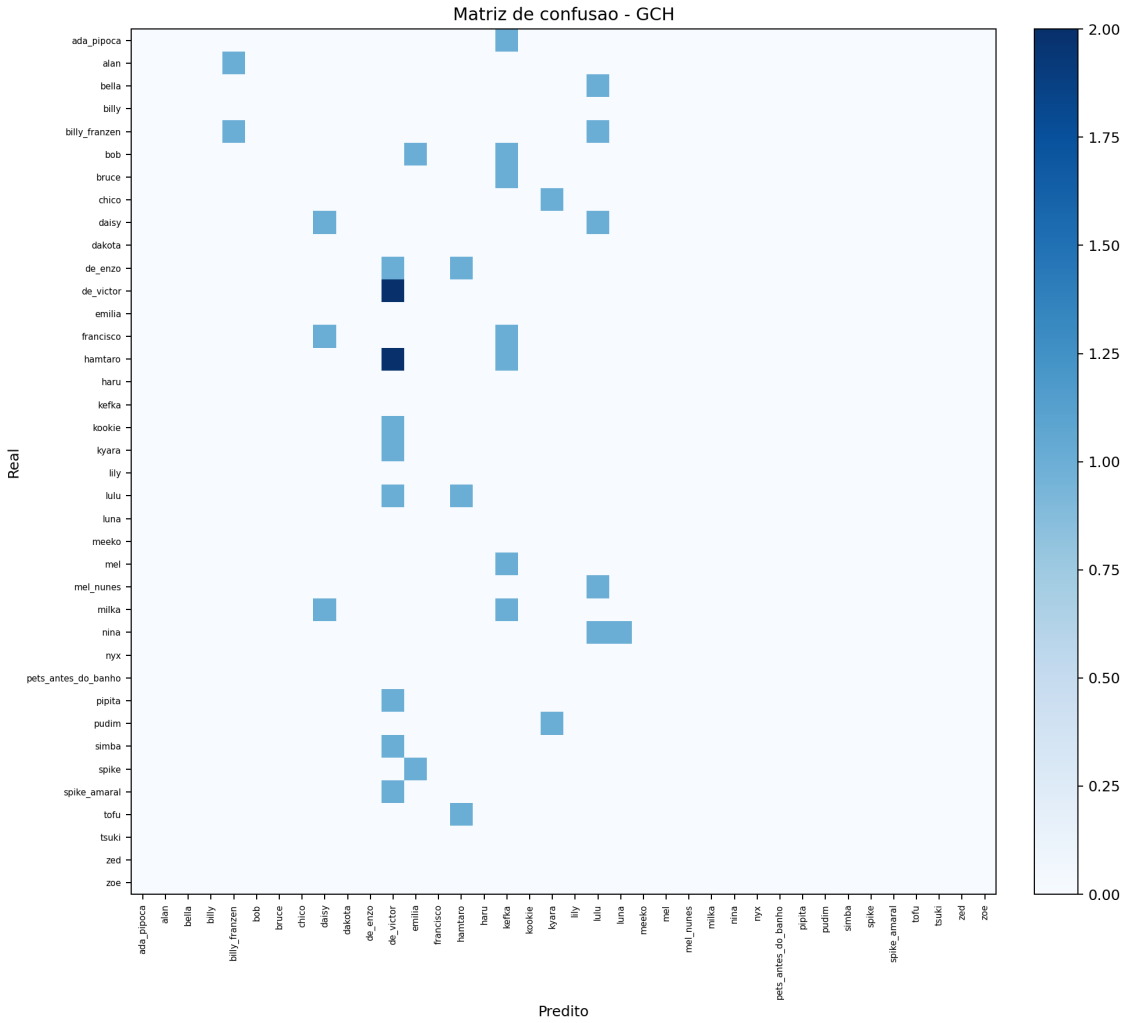


Figura 4: Matriz de confusão para o descritor GCH, usado como referência simples de cor.

As matrizes de confusão mostram que a tarefa é difícil porque o conjunto de teste tem apenas 36 imagens distribuídas em muitas classes. Mesmo nos melhores métodos, várias classes possuem apenas uma imagem no conjunto de teste, de modo que poucos erros já deslocam fortemente a acurácia por classe. Essa característica também torna a comparação entre descritores sensível ao split: diferenças de uma ou duas imagens classificadas corretamente já alteram a acurácia em vários pontos percentuais. Ainda assim, a matriz da combinação de todos os descritores apresenta mais acertos na diagonal do que a referência GCH, indicando que a união de múltiplas semânticas visuais ajuda a reduzir algumas confusões. A matriz do GCH, por sua vez, evidencia a limitação de usar apenas distribuição global de cor: pets visualmente parecidos ou fotografados em fundos semelhantes tendem a ser confundidos. Para manter a leitura principal concentrada nos resultados mais informativos, as demais matrizes aparecem reunidas na seção de matrizes de confusão por método.

4.2 Busca

A Tabela 3 mostra os resultados de busca por similaridade. A melhor configuração combinou GCH, LBP e GLCM, com mAP de 0.1346 e P@1 de 0.2970. A combinação que

também inclui Gabor teve desempenho muito próximo, mas ligeiramente inferior em mAP.

Tabela 3: Resultados de busca por distância euclidiana.

Método	Dim.	mAP	P@1	P@5	P@10
GCH	256	0.0997	0.2262	0.1248	0.0940
LBP	256	0.1183	0.2643	0.1439	0.1125
GLCM	10	0.0877	0.1635	0.1003	0.0801
HOG	1764	0.0837	0.1662	0.0926	0.0706
Correlograma	48	0.0952	0.1989	0.1117	0.0888
Gabor	120	0.0973	0.1907	0.1188	0.0926
GCH+LBP	512	0.1327	0.2807	0.1619	0.1294
GCH+GLCM	266	0.1082	0.2534	0.1390	0.1025
GCH+LBP+GLCM	522	0.1346	0.2970	0.1640	0.1278
GCH+LBP+GLCM+Gabor	642	0.1329	0.2888	0.1668	0.1283
Todos	2454	0.1173	0.2561	0.1422	0.1093

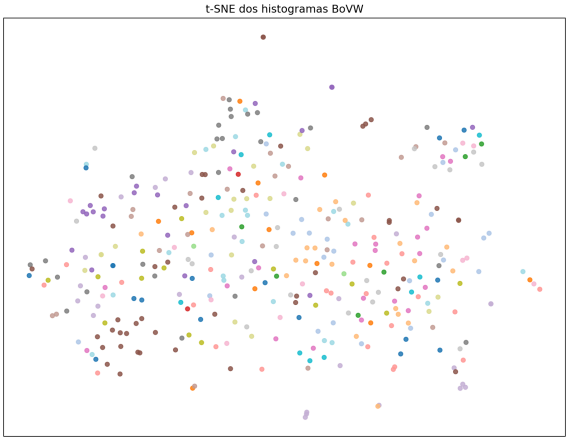
4.3 Bag of Visual Words

O BoVW foi implementado com descritores locais densos calculados diretamente a partir de gradientes. A imagem foi dividida em patches de 16x16 pixels com passo 16. Para cada patch, foi extraído um histograma local de orientações de gradiente, complementado por média e desvio padrão de intensidade. Os descritores locais foram agrupados por K-Means com $K = 200$, e cada imagem foi representada por um histograma normalizado de visual words. O vocabulário visual foi aprendido de forma não supervisionada sobre a coleção completa usada na busca, seguindo a mesma lógica da avaliação de retrieval: todas as imagens participam da coleção consultável, mas os rótulos não são usados para construir o dicionário.

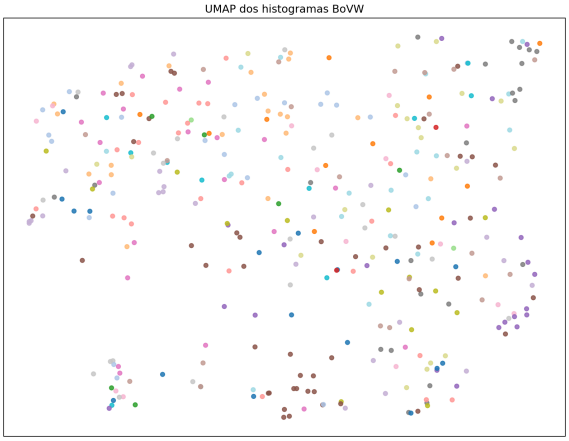
Tabela 4: Resultados de busca com BoVW.

Método	K	Dim.	mAP	P@1	P@5	P@10
BoVW	200	200	0.1073	0.2262	0.1297	0.1063

Comparado aos descritores globais e semiglobais, o BoVW teve desempenho intermediário. Seu mAP de 0.1073 supera descritores isolados como GLCM, HOG, Correlograma, Gabor e GCH, mas fica abaixo de LBP e das melhores combinações, especialmente GCH+LBP+GLCM. Portanto, o BoVW capturou estrutura visual útil, mas não foi a representação mais discriminante para esta base.



(a) t-SNE com todas as classes.



(b) UMAP com todas as classes.

Figura 5: Projeções 2D dos histogramas BoVW.

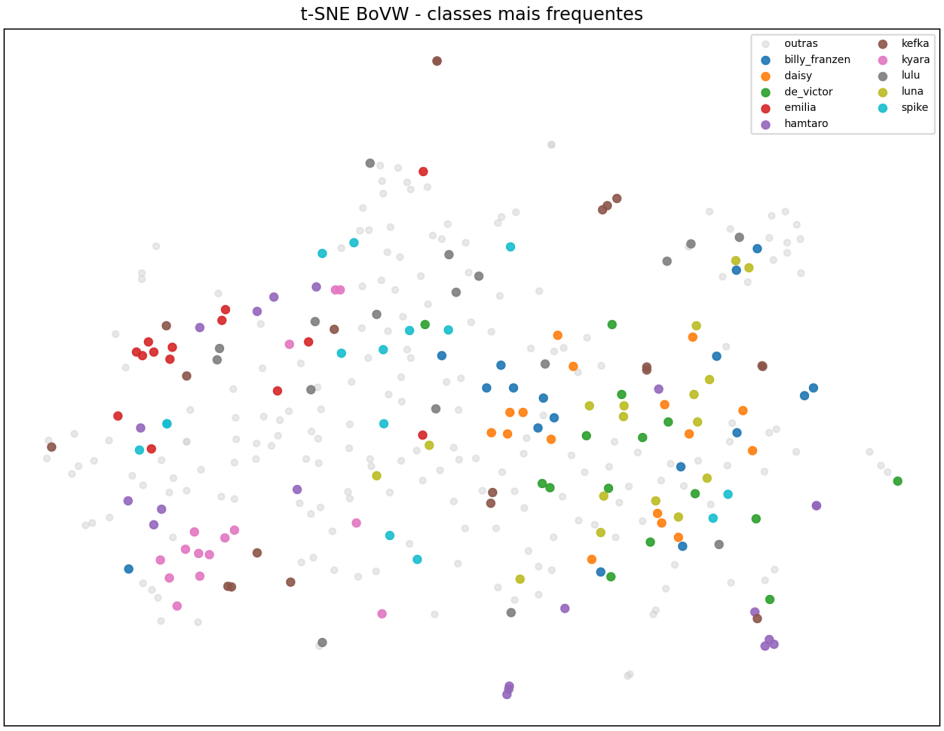


Figura 6: t-SNE dos histogramas BoVW destacando as classes mais frequentes.

5 Análise por Descritor

5.1 GCH

O GCH confirmou parcialmente sua hipótese inicial. Ele obteve apenas 11,11% de acurácia no teste de classificação, mas alcançou mAP de 0.0997 na busca. Isso indica que a cor isolada não foi suficiente para treinar um classificador robusto no split 80/10/10, mas ainda forneceu alguma informação útil para vizinhança visual. A P@1 de 0.2262 mostra que, em uma fração relevante das consultas, a imagem mais próxima tem o mesmo pet,

embora a representação global ainda confunda casos em que fundo e iluminação têm forte peso no histograma.

Nas combinações, o GCH foi importante. Quando combinado com LBP e GLCM, produziu a melhor busca. Isso sugere que a cor ajuda a formar um primeiro agrupamento visual, enquanto textura refina a distinção entre imagens de cores próximas.

5.2 LBP

O LBP teve acurácia de teste de 16,67% na classificação, apesar de ter obtido a maior acurácia de validação entre os descritores individuais. Essa diferença sugere instabilidade causada pelo pequeno número de imagens por classe no conjunto de validação e teste. Na busca, porém, o LBP foi o melhor descritor isolado, com mAP de 0.1183. Uma explicação provável é que a textura da pelagem, junto com padrões locais de fundo e iluminação, cria uma assinatura visual útil para ranqueamento, mesmo que não seja suficiente para classificação supervisionada robusta.

A hipótese de complementaridade com cor também se confirmou. GCH+LBP melhorou a busca em relação a ambos isoladamente, chegando a mAP de 0.1327. A combinação com GLCM melhorou ainda mais, indicando que textura local e textura estatística capturam aspectos diferentes da imagem.

5.3 GLCM/Haralick

O GLCM teve desempenho baixo isoladamente: 11,11% de acurácia e mAP de 0.0877. Isso mostra que suas estatísticas globais de textura são compactas demais para separar muitas classes individuais. Ainda assim, o descritor foi útil em algumas combinações, principalmente na busca, onde GCH+LBP+GLCM foi o melhor método.

O ganho observado em GCH+LBP+GLCM indica que o GLCM adiciona informação complementar. Mesmo com apenas 10 dimensões, ele melhora a combinação de cor e textura local, provavelmente por capturar regularidade e contraste em escala mais global.

5.4 HOG

Neste split de dados, o HOG obteve a maior acurácia individual na classificação, com 27,78% de acurácia no teste, mas teve o pior mAP na busca, com 0.0837. Esse contraste mostra que a informação de forma e bordas pode ser explorada pela SVM no conjunto supervisionado, mas não define um bom espaço de vizinhança euclidiana para retrieval. A hipótese de fragilidade por pose continua válida para a busca, pois a base não possui enquadramento controlado: o mesmo pet pode aparecer deitado, em pé, parcialmente cortado ou com fundo muito diferente.

Além disso, o HOG tem 1764 dimensões. Na classificação, sua inclusão contribuiu para que a combinação de todos os descritores alcançasse o melhor resultado. Na busca, entretanto, a combinação de todos os descritores ficou abaixo de combinações mais compactas, sugerindo que a mesma alta dimensionalidade que ajuda o classificador pode prejudicar a distância euclidiana.

5.5 Correlograma de Cores

O correlograma obteve 19,44% de acurácia no teste e mAP de 0.0952. O resultado superou o GCH na classificação, mas ficou próximo dele na busca. A hipótese de que informação

espacial de cor poderia ajudar se confirmou apenas parcialmente.

Uma possível causa é a decisão de usar apenas o canal de matiz. Essa escolha torna o vetor compacto, mas remove informação de saturação e brilho. Além disso, padrões espaciais de cor mudam bastante com pose, recorte e fundo. Assim, o ganho de espacialidade não compensou a perda de informação cromática completa.

5.6 Gabor

O Gabor teve acurácia individual de 22,22% e mAP de 0.0973. Embora não tenha sido o melhor descritor isolado, ele melhorou a classificação quando combinado com GCH, LBP e GLCM: essa combinação alcançou 30,56% de acurácia, acima de GCH+LBP+GLCM.

Isso indica que filtros de Gabor capturam textura orientada complementar. O descritor não foi dominante sozinho, mas acrescentou informação útil quando combinado com cor e outras texturas. Na busca, porém, sua adição não melhorou o mAP em relação a GCH+LBP+GLCM, sugerindo que a distância euclidiana no espaço combinado é mais sensível à inclusão de blocos adicionais.

6 Análise Comparativa e Estudos de Caso da Busca por Similaridade

Os resultados mostram uma diferença importante entre classificação e busca. Na classificação, a melhor configuração foi a combinação de todos os descritores, sugerindo que a SVM conseguiu aproveitar informação complementar de cor, textura e forma. Na busca, a melhor combinação foi GCH+LBP+GLCM, sem HOG, Correlograma ou Gabor. Isso indica que o melhor espaço para um classificador supervisionado não é necessariamente o melhor espaço para vizinhança euclidiana.

Outro resultado relevante é que a combinação de todos os descritores foi a melhor na classificação, com 38,89% de acurácia, mas não foi a melhor na busca, com mAP de 0.1173. Esse comportamento mostra que alta dimensionalidade e múltiplas semânticas podem beneficiar o classificador supervisionado, mas prejudicar o ranqueamento por distância euclidiana quando há descritores pouco alinhados à noção de similaridade visual.

Figura 7: Estudo de caso bem-sucedido da busca por similaridade com GCH+LBP+GLCM para a consulta da classe `kyara`.

Figura 8: Segundo estudo de caso bem-sucedido da busca por similaridade com GCH+LBP+GLCM para a consulta da classe `nina`.

Figura 9: Estudo de caso de falha da busca por similaridade com GCH+LBP+GLCM para a consulta da classe `simba`.

No primeiro caso de sucesso, a consulta de `kyara` retorna várias imagens do mesmo pet nas primeiras posições. A combinação GCH+LBP+GLCM parece aproveitar simultaneamente a cor predominante, a textura da pelagem e o contexto visual semelhante das imagens. Os erros restantes aparecem quando outros animais ou fundos compartilham tons e padrões locais próximos.

No segundo caso, a consulta de `nina` também mostra bom comportamento: os primeiros resultados preservam a pelagem clara, o enquadramento do corpo e a textura do ambiente externo. A presença de resultados incorretos, como `tofu` e `bob`, mostra a limitação do método: imagens de animais diferentes podem ficar próximas quando têm cores claras, pose parecida ou textura de fundo semelhante.

No caso de falha com `simba`, nenhum dos primeiros resultados pertence à classe correta. A consulta é um gato com textura listrada e fundo escuro, mas os descritores aproximaram imagens com padrões locais, tons ou fundos visualmente parecidos. Esse estudo de caso evidencia que os descritores não reconhecem a identidade do animal; eles apenas comparam estatísticas visuais de baixo nível.

7 Hipóteses sobre a Distribuição BoVW

As projeções BoVW mostram uma distribuição com sobreposição considerável entre classes. Isso é coerente com o mAP de 0.1073, que fica abaixo da melhor combinação global. Como as visual words foram construídas a partir de patches densos de gradiente, elas descrevem padrões locais de orientação e intensidade, mas não incorporam cor diretamente. Essa limitação reduz a capacidade de separar pets cuja identidade depende muito da pelagem e da coloração.

Mesmo assim, a projeção não é completamente aleatória. Algumas regiões mostram pequenos agrupamentos, o que sugere que o BoVW captura recorrências locais de textura, pose ou fundo. Na visualização com classes mais frequentes, é possível observar que algumas classes aparecem mais concentradas do que outras. Uma hipótese é que pets com muitas fotos em ambientes semelhantes geram histogramas de visual words mais consistentes, enquanto pets fotografados em poses e fundos variados se espalham mais.

Comparando t-SNE e UMAP, as duas projeções reforçam a mesma conclusão: há estrutura local, mas não há separação limpa entre a maioria dos pets. Portanto, o BoVW implementado é útil como representação adicional e atende ao objetivo experimental, mas não supera as melhores combinações de descritores globais.

8 Conclusão

O trabalho implementou e avaliou um sistema completo de busca e classificação de animais de estimação usando descritores clássicos. Foram extraídos seis descritores para todas as imagens, incluindo Gabor como descritor não visto em aula. Os descritores foram avaliados individualmente e em combinações nas tarefas de classificação e busca, e também foi implementado um método Bag of Visual Words com visualizações em t-SNE e UMAP.

A primeira pergunta experimental era quais descritores capturam melhor diferenças entre pets individuais. Os resultados dependem da tarefa: HOG foi o melhor descritor individual para classificação, enquanto LBP foi o melhor descritor individual para busca. Assim, forma e bordas ajudaram mais a SVM, enquanto textura local foi mais adequada para ranqueamento por distância euclidiana. A hipótese de que HOG seria sensível a pose e enquadramento se confirmou na busca, onde ele teve o pior mAP.

A segunda pergunta era se combinações seriam mais robustas que descritores isolados. A resposta é positiva, mas depende da tarefa. Na classificação, a melhor configuração foi a combinação de todos os descritores, com 38,89% de acurácia no teste. Na busca, a melhor configuração foi GCH+LBP+GLCM, com mAP de 0.1346 e P@1 de 0.2970. Portanto, as melhores combinações superaram os descritores isolados em alguns cenários, mas a combinação ideal para SVM não é a mesma para distância euclidiana.

A terceira pergunta era se BoVW conseguiria competir com os descritores globais. O BoVW obteve mAP de 0.1073 e P@1 de 0.2262. Esse resultado é competitivo com alguns descritores isolados, mas fica abaixo das melhores combinações globais. As visualizações confirmam uma distribuição com agrupamentos locais, mas forte sobreposição entre classes, validando a hipótese de que as visual words capturam alguma estrutura, mas não separam claramente a identidade dos pets.

Como limitações, a base possui poucas imagens por classe e grande variação de captura. Além disso, descritores clássicos comparam propriedades visuais de baixo nível e não reconhecem semanticamente o pet. Como trabalhos futuros, seria possível avaliar redução de dimensionalidade antes da classificação, testar descritores locais mais ricos e comparar os resultados com características extraídas de redes neurais pré-treinadas.

9 Matrizes de Confusão por Método

Esta seção reúne as matrizes de confusão completas dos métodos avaliados na classificação. As figuras complementam a Tabela 2: a tabela resume os padrões principais, enquanto as matrizes abaixo permitem verificar a distribuição completa de acertos e erros por classe.

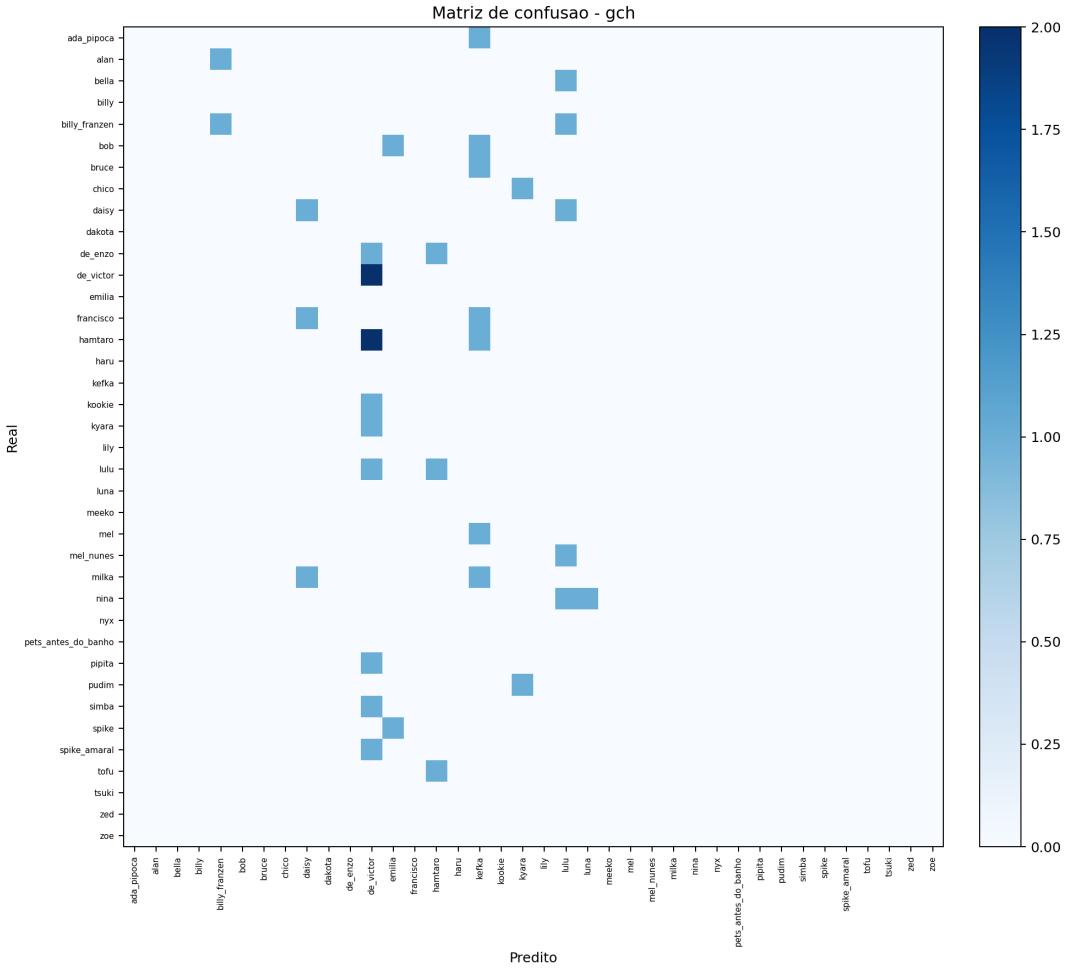


Figura 10: Matriz de confusão completa para GCH.

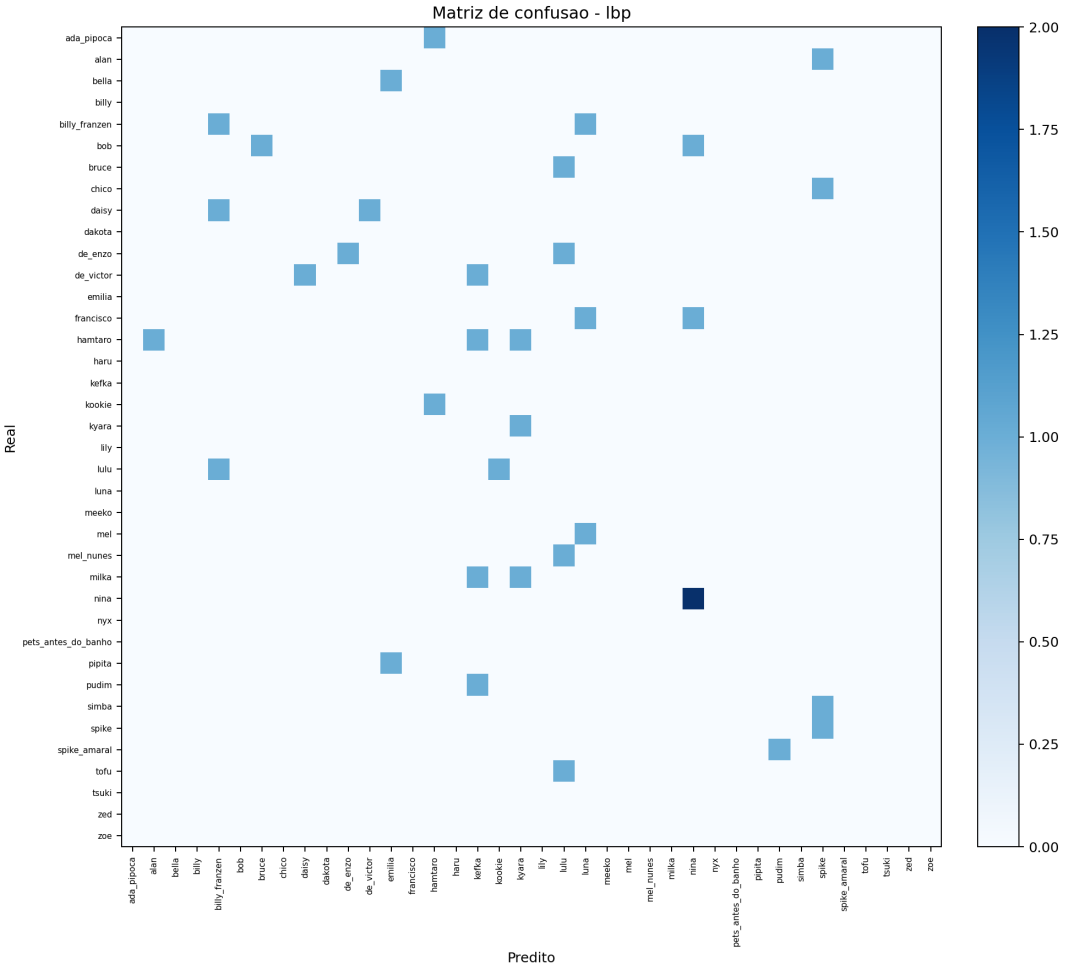


Figura 11: Matriz de confusão completa para LBP.

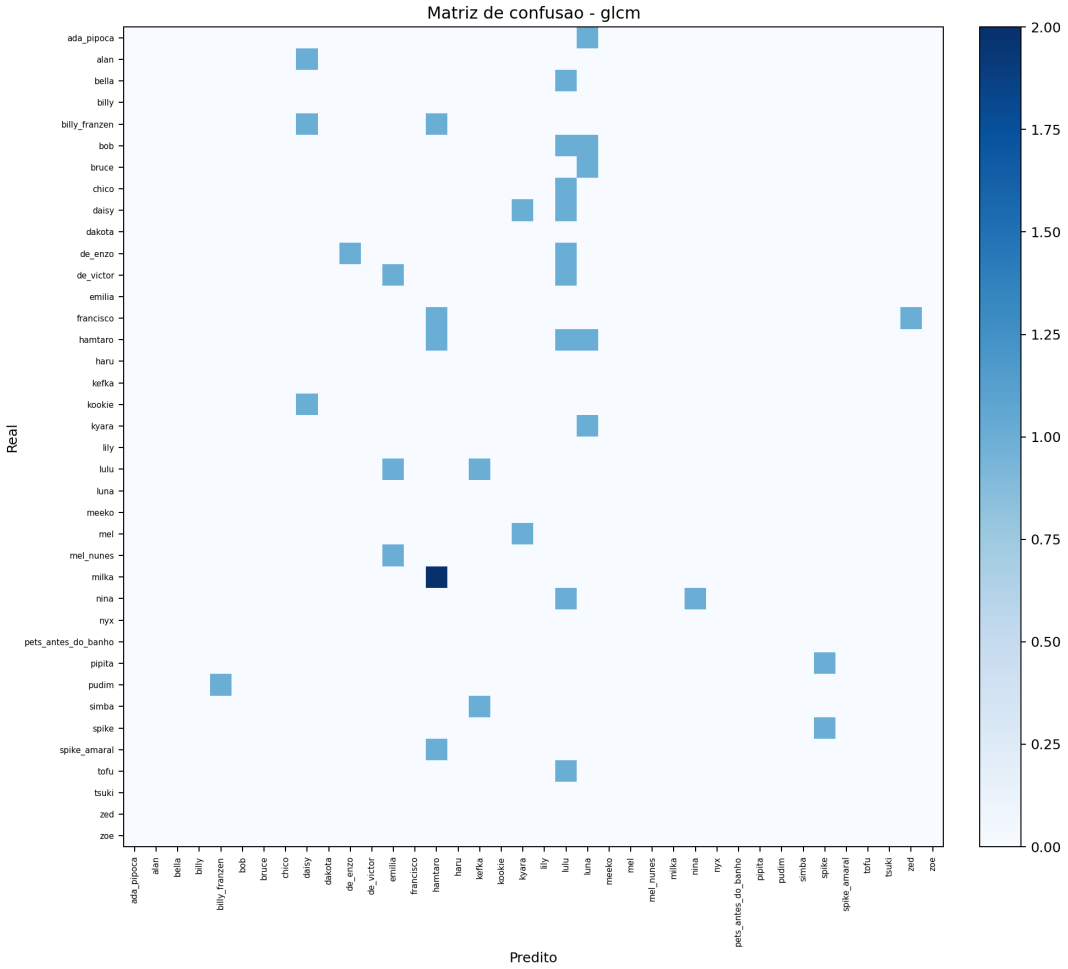


Figura 12: Matriz de confusão completa para GLCM.

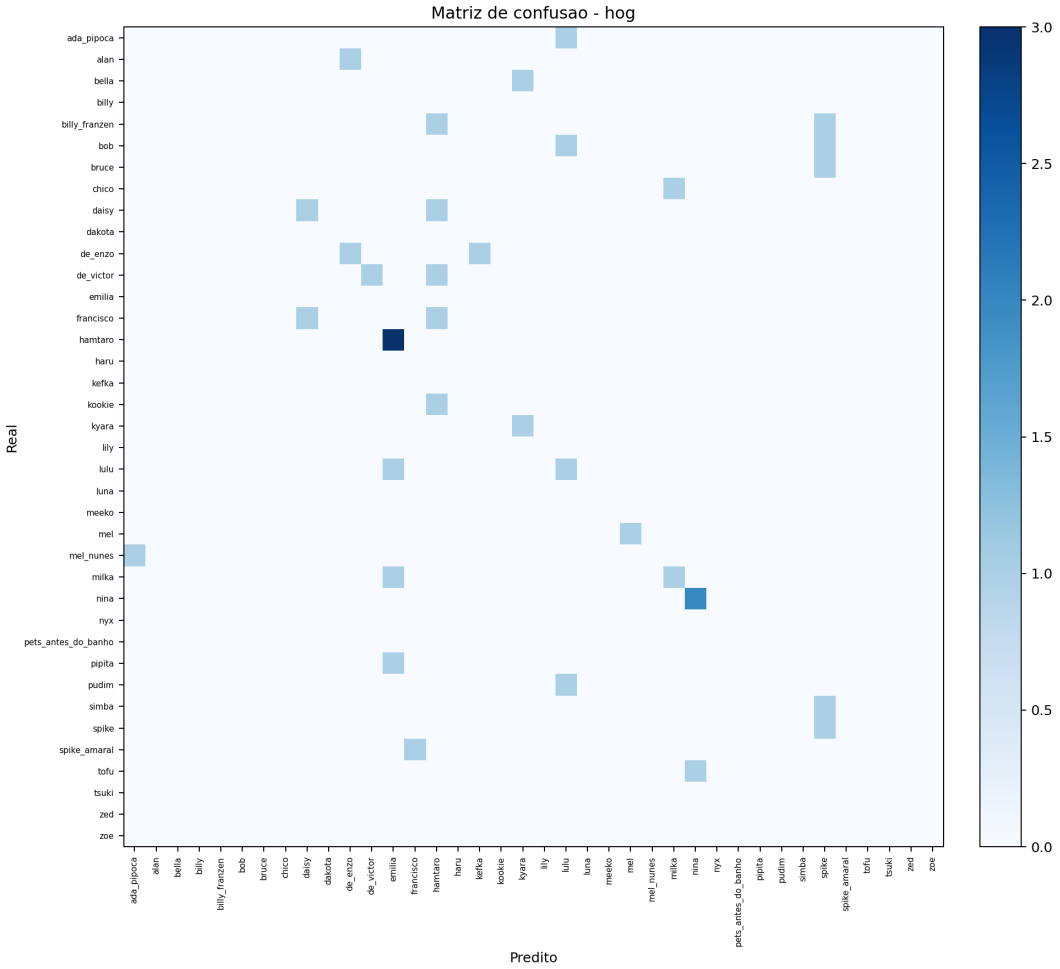


Figura 13: Matriz de confusão completa para HOG.

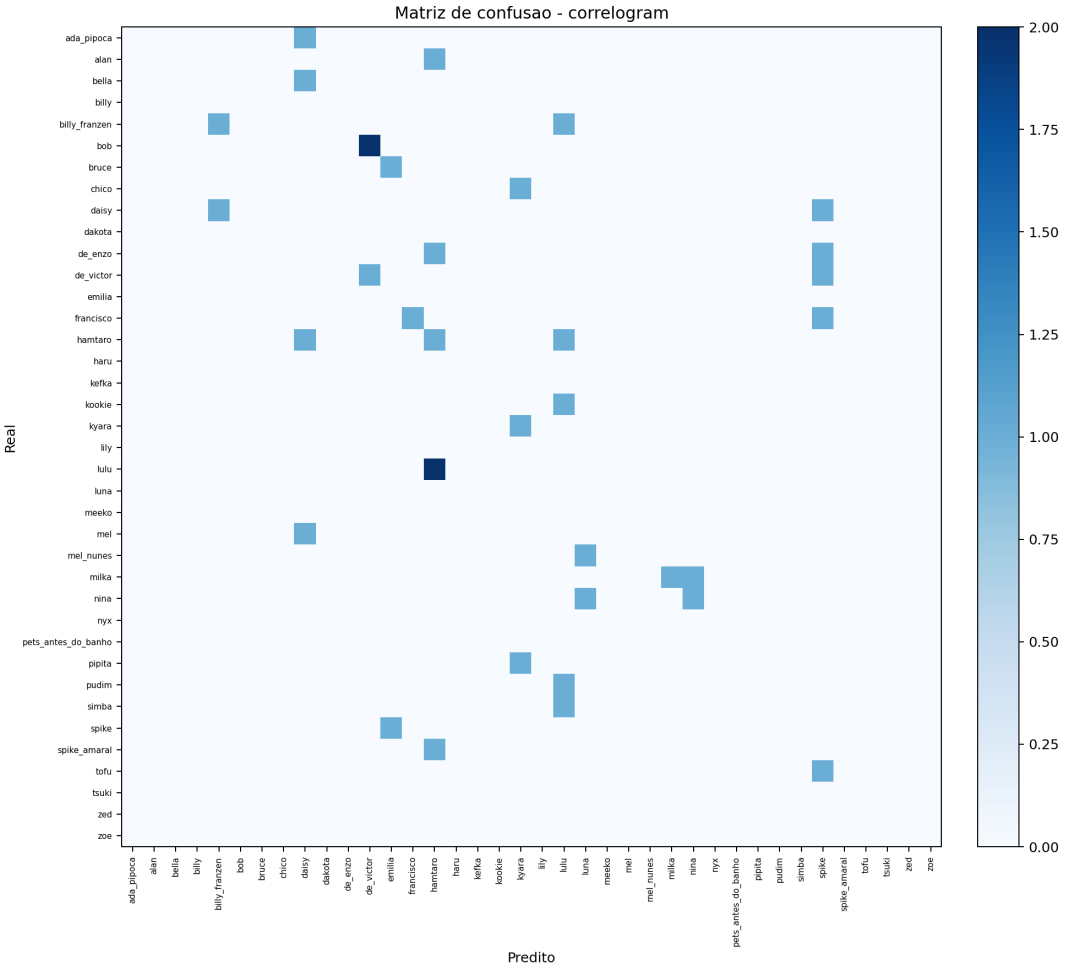


Figura 14: Matriz de confusão completa para Correlograma.

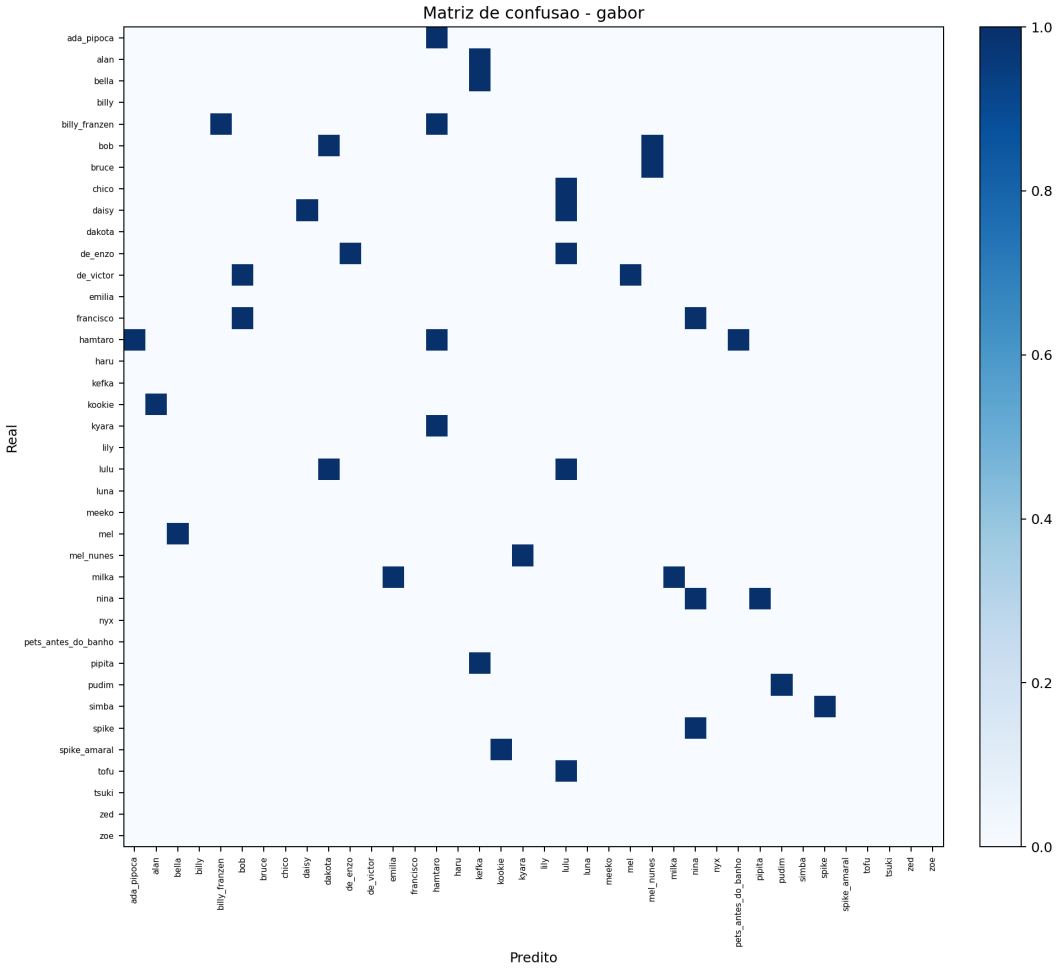


Figura 15: Matriz de confusão completa para Gabor.

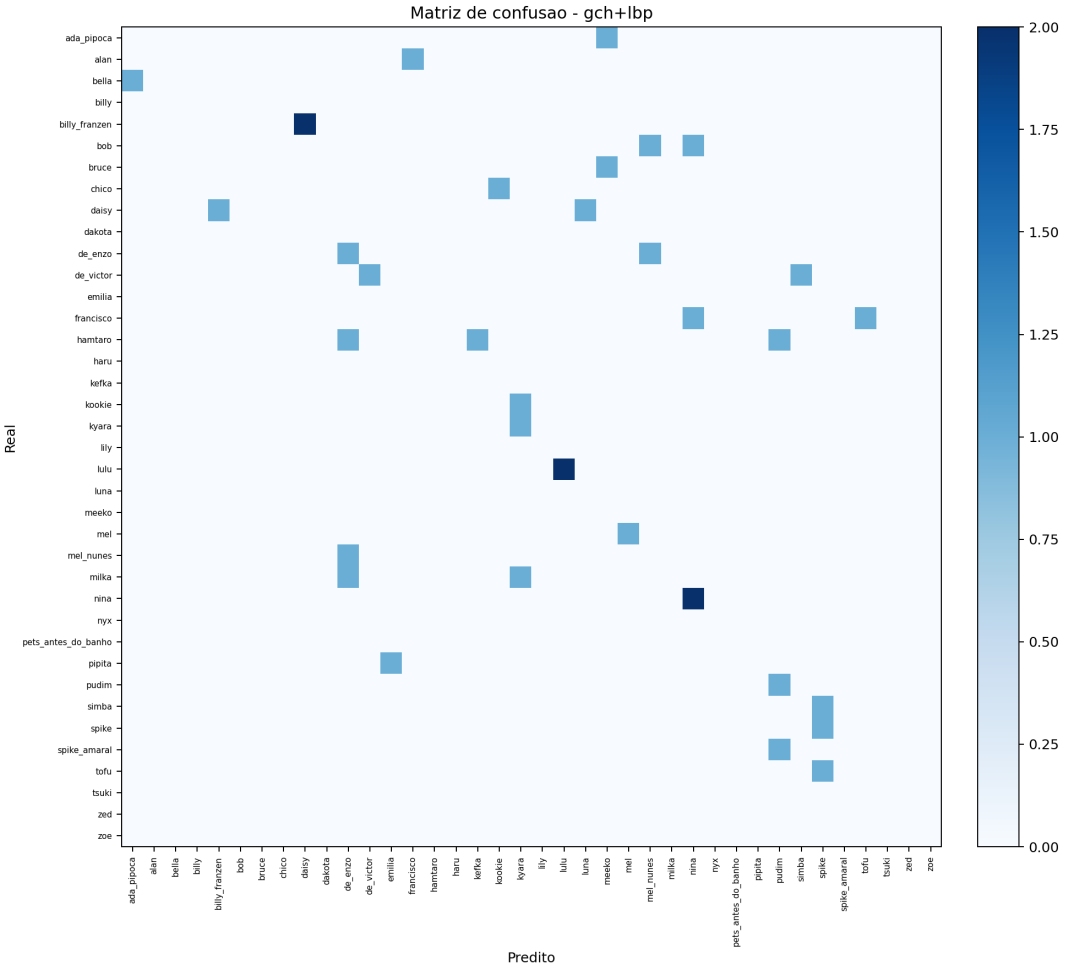


Figura 16: Matriz de confusão completa para GCH+LBP.

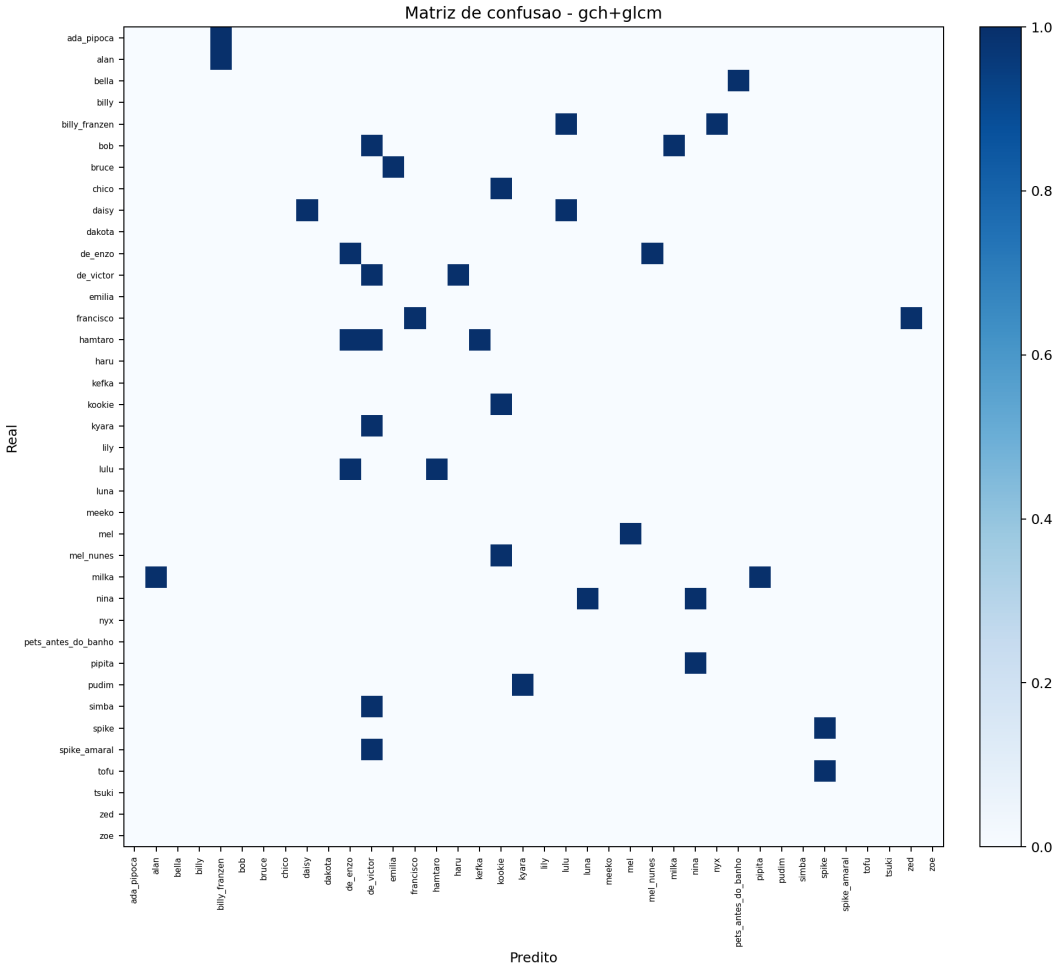


Figura 17: Matriz de confusão completa para GCH+GLCM.

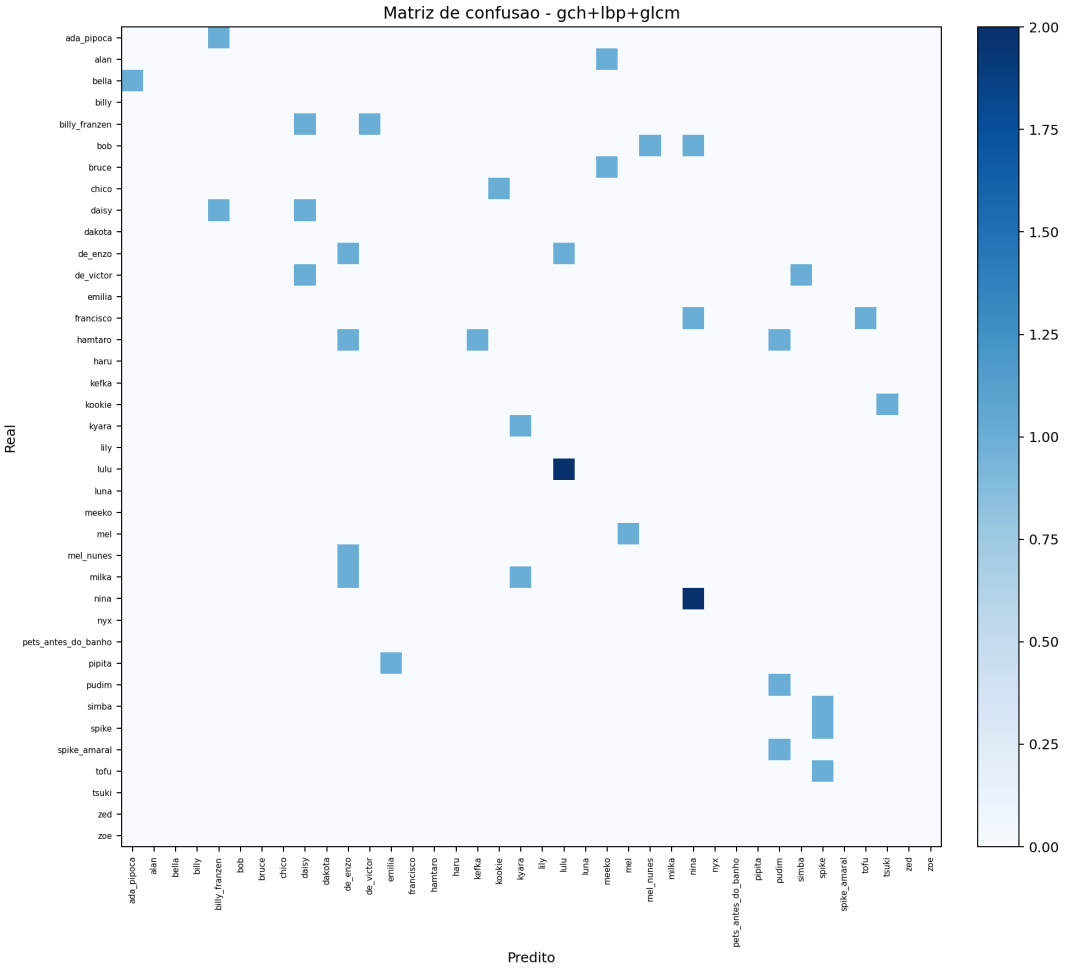


Figura 18: Matriz de confusão completa para GCH+LBP+GLCM.

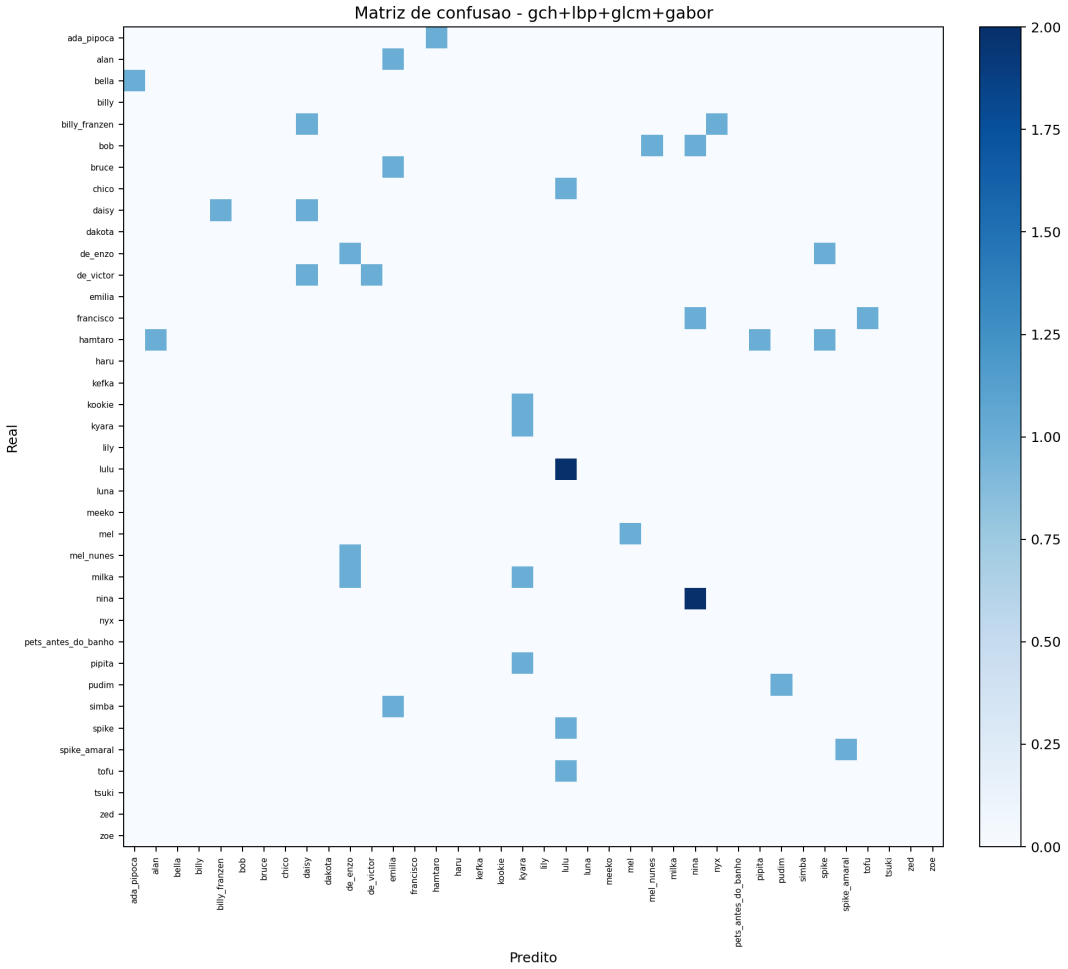


Figura 19: Matriz de confusão completa para GCH+LBP+GLCM+Gabor.

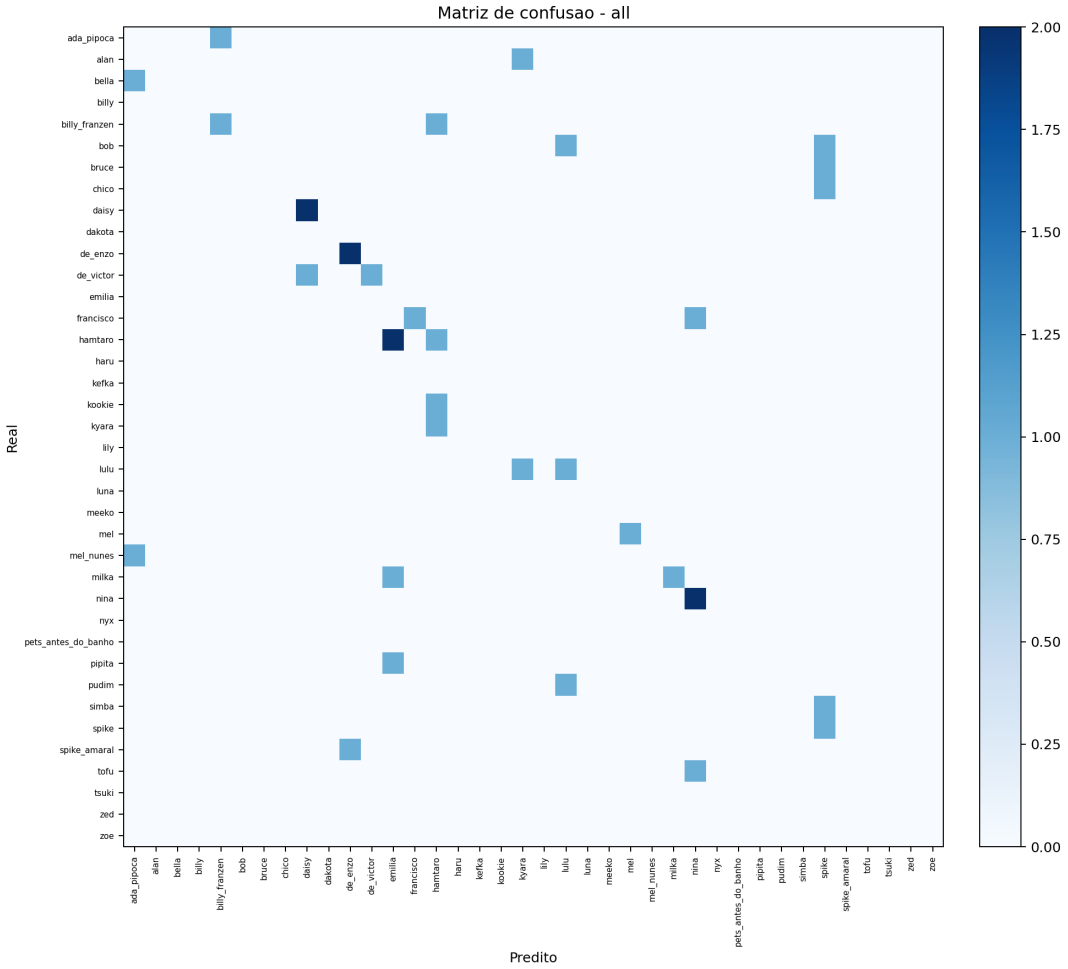


Figura 20: Matriz de confusão completa para a combinação de todos os descritores.