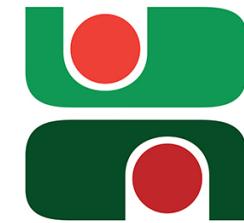
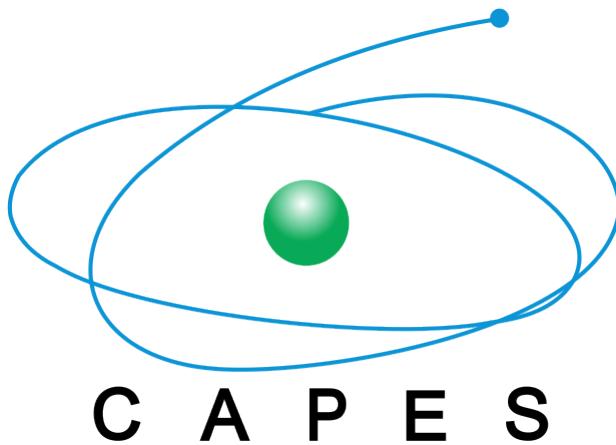


# Reconhecimento de Produtos por Imagem Utilizando Palavras Visuais e Redes Neurais Convolucionais

Guilherme Defreitas Juraszek

Orientador: Dr. Alexandre Gonçalves Silva  
Coorientador: Dr. André Tavares da Silva

# Agradecimentos



# Agenda

- Motivação
- Objetivos
- Trabalhos Relacionados
- Conceitos
- Experimentos
- Resultados
- Conclusão

# Motivação

- Reconhecimento de produtos utilizando apenas imagens 2D, sem qualquer informação adicional como descrição, código de barras ou contexto.
- Reconhecimento consiste em determinar a qual categoria um produto pertence.
- A definição das categorias acontece no momento do treinamento.

# Motivação

- Um objeto tridimensional pode ser representado por meio de uma grande quantidade de imagens bidimensionais (Pinto et. al. 2008)
  - Posição
  - Orientação
  - Tamanho
  - Iluminação
  - Contexto
  - Variações intra-classes
  - Oclusões
  - Deformações



Exemplo de variação intra-classe com objetos pertencentes a uma mesma categoria (cafeteiras).

Fonte: Produção do próprio autor.

# Objetivos

- Comparar as técnicas Conjunto de Palavras Visuais (*Bag of Visual Words* – BOVW), Redes Neurais Convolucionais (CNN) e CNN como um descritor.
- Identificar qual das três técnicas (BOVW, CNN e CNN como descritor natural) oferece melhor acurácia.
- Identificar qual das três técnicas oferece menor tempo de processamento.
- Identificar qual das três técnicas oferece a melhor acurácia em bases de diferentes tamanhos.

# Objetivos

- Identificar qual o melhor descritor, entre SIFT, SURF e MSER, na técnica de BOVW.
- Identificar como o tamanho da imagem influencia na acurácia do reconhecimento.
- Identificar como o tamanho do dicionário de palavras visuais influencia na técnica BOVW.

# Trabalhos Relacionados

Trabalhos divididos em duas categorias

- Recuperação de imagens por conteúdo (*Content-Based Image Retrieval – CBIR*)
  - Extração e comparação de descritores utilizando uma função de distância.
- Reconhecimento ou classificação de objetos em uma imagem.
  - Aprendizado de máquina

# Trabalhos Relacionados

- Torres e Falção (2006) demonstram o processo comum em sistemas CBIR através da extração de características como forma, textura e distribuição das cores.
- Datta et al. (2008) – Estudo de 300 trabalhos e contribuições em sistemas CBIR onde destaca o crescimento da utilização de técnicas de aprendizado de máquina combinados com descritores artificiais.

# Trabalhos Relacionados

Jarret et al. (2009) classificam sistemas de reconhecimento de imagens em duas categorias:

- Algoritmos de estágio único
  - Apenas uma etapa de extração de características
    - Descritores artificiais, fórmulas fixas independente do contexto ou da base de imagens.
    - Conjunto de palavras visuais
- Algoritmos de dois ou mais estágios
  - Duas ou mais etapas organizadas de forma hierárquica onde cada etapa identifica características simples até características mais elaboradas.
    - Bordas -> Cantos -> Formas -> Objetos -> Cenas/Ações
    - Concreto -> Abstrato
  - Normalmente o classificador utilizado aprende o descritor na etapa de treinamento.

# Trabalhos Relacionados

- Algoritmos de estágio único
  - *Pyramid Match Kernel* PMK-SVM (Lazebnik et al. 2006)
  - OPF não supervisionado para agrupamento (Afonso et al. 2002).
  - Li e Perona (2005) mostram que a extração de característica de forma densa obteve um melhor resultado do que pontos de interesse para classificação de cenas.
- Algoritmos de dois ou mais estágios
  - Jarret et al (2009) mostram um algoritmo de dois estágios capaz de aprender uma representação interna (descritor) e obter um resultado comparável à técnica PMK-SVM.

# Trabalhos Relacionados

- Algoritmos de dois ou mais estágios
  - Cirezan et al. (2012) Obtiveram **18,5%** de erro no reconhecimento de objetos na base CIFAR-10 utilizando uma rede neural convolucional (CNN).
    - CNN + Regularização Dropout + aumento artificial de dados = **16,6%** de erro na CIFAR-10 (Hinton et al, 2012).
    - CNN + Stochastic Pooling = **15,12%** de erro na CIFAR-10 (Zeiler e Fergus, 2013)
    - CNN + SVM = **11,9%** de erro na CIFAR-10 (Tang, 2013)
    - Maxout + Dropout = **11,68%** de erro na CIFAR-10 sem aumento de dados e **9,35%** com aumento (Goodfellow, 2013)

# Trabalhos Relacionados

- Algoritmos de dois ou mais estágios
  - Aplicação de uma CNN com filtros treinados de forma não supervisionada (Kavukcuoglu et al. 2012).
  - Sermanet et al. (2013) apresentam uma CNN (*Overfeat*) que vence o concurso *ImageNet Large Scale Visual Recognition Challenge 2013* (ILSVRC2013) em uma base de 1,2 milhão de imagens divididas em 1000 categorias.
  - Razavian et al. (2014) utilizam a *Overfeat* para extração de um vetor de características classificador com SVM mostrando resultados compatíveis com o estado da arte.

# Trabalhos Relacionados

Utilizar apenas o valor de intensidade dos pixels não oferecem um bom resultado quando combinados com um classificador linear (Wallraven et al. 2003).

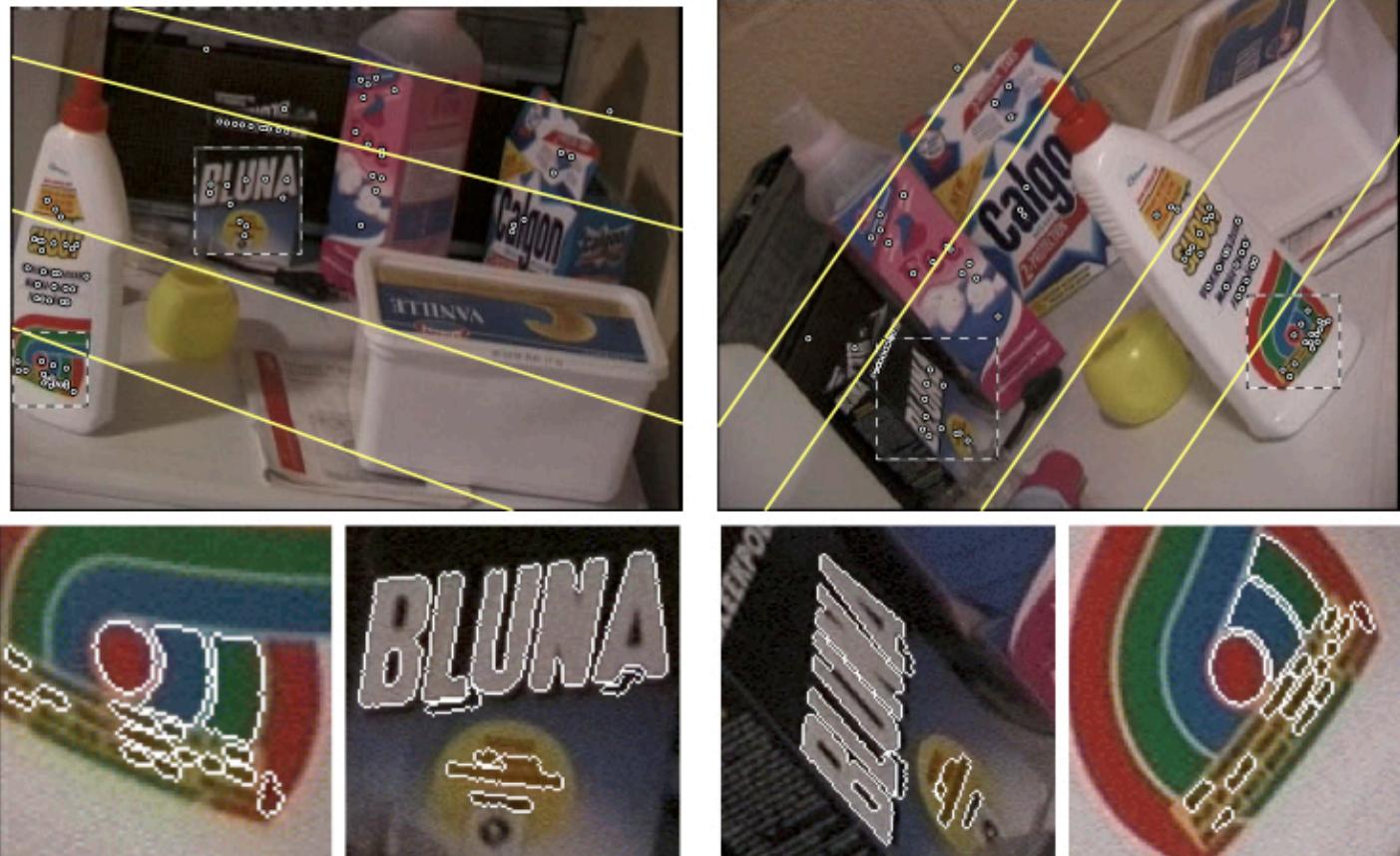
Alguns descritores estudados:

- SIFT (Lowe, 2004)
- HOG (Dalal e Triggs, 2005)
- FAST (Rosten e Drummond, 2006)
- ORB (Rublee et al. 2011)
- MSER (Matas et al. 2002)
- SURF (Bay et. Al. 2008)

# Descritores Locais

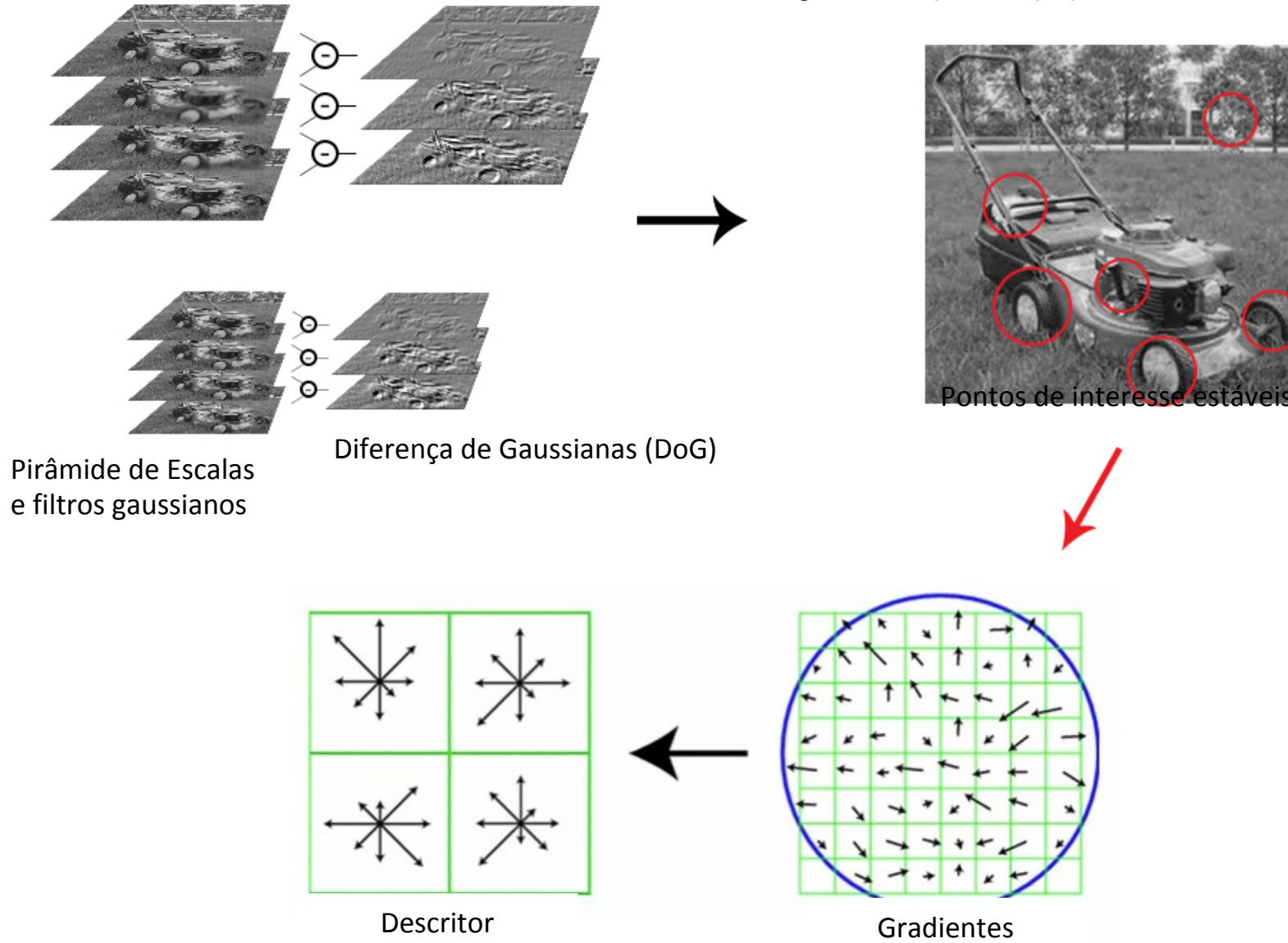
- Fornecer informações sobre uma imagem capaz de identificar os mesmos pontos em duas imagens diferentes do mesmo objeto.
- Um descritor deve extrair características invariantes à iluminação, rotação, escala, translação e deformações de diferentes ângulos.
- Duas etapas
  - Identificar os pontos de interesse
  - Descrever os pontos de interesse

# Descritores



Exemplos de regiões estáveis detectadas em diferentes ângulos.  
(Matas et al. 2002)

# *Scale Invariant Feature Transform (SIFT) (Lowe, 2004)*



Processo de localização e descrição do SIFT.  
Fonte: Produção do próprio autor.

# *Speed-Up Robust Features* (SURF) (Bay et al. 2008)

- Inspirado no SIFT
- Objetivo de ser mais rápido na extração e comparação de descritores.
- Pontos de interesse encontrados com base em grandes variações de intensidade.

# *Maximally Stable Extremal Regions*

## (MSER) (Matas et al. 2002)

- Algoritmo de localização de pontos de interesse.
  - Não fornece um descritor
- Objetivo de localizar pontos de interesse robustos diante de mudanças de perspectiva.
- Aplica diferentes limiares para detectar bordas com grande variação de intensidade.
- Apresenta semelhanças com o algoritmo Watershed (Beucher e Lantuéjoul, 1979).
  - Watershed busca encontrar onde componentes conexos se juntam
  - MSER busca o conjunto de limiares onde um grupo de componentes conexos permanece estável.

# Agrupamento e Classificação

KMeans

OPF

SVM

# KMeans (Macqueen, 1967)

- Particiona um conjunto de  $n$  observações (neste trabalho, os descritores) em um conjunto de grupos de tamanho  $k$  de forma que cada observação pertença ao grupo mais próximo, de acordo com uma função de distância.

# Floresta de Caminhos Ótimos (Optimum-path Forest - OPF) (Papa et al. 2009)

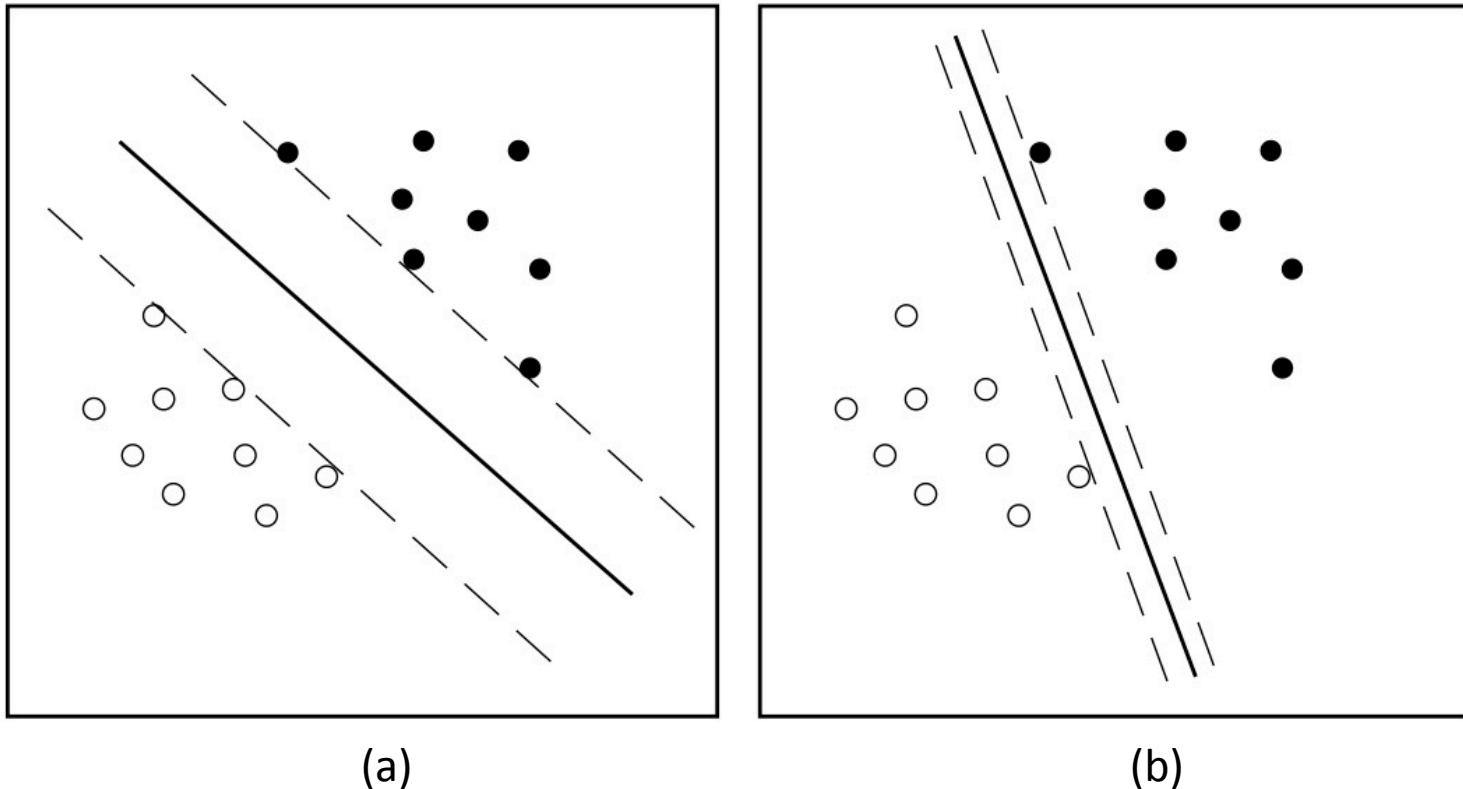
- Algoritmo de classificação com base em grafos.
- Etapas
  - Escolha dos protótipos
    - Pelo menos um nó por classe.
    - Aleatória ou com base em alguma heurística (ex: densidade)
  - O processo de conquista ocorre através da comparação do nó protótipo e do nó a ser analisado de acordo com uma função de distância.
- Após o treinamento o classificador é uma floresta de caminhos ótimos onde cada exemplo pertence a uma árvore com uma raiz protótipo.

# Máquinas de Vetor de Suporte (Support Vector Machines – SVM)

## (Cortes e Vapnik, 1995)

- Algoritmo de classificação
- Busca a divisão do hiperplano maximizando a margem entre as classes
- É um separador linear
  - Em problemas não linearmente separáveis é necessário aplicar uma função de transformação (*kernel*) para possibilitar a separação linear.
    - Poligonal, gaussiano ou radial, sigmoid entre outros.

# Máquinas de Vetor de Suporte (Support Vector Machines – SVM) (Cortes e Vapnik, 1995)

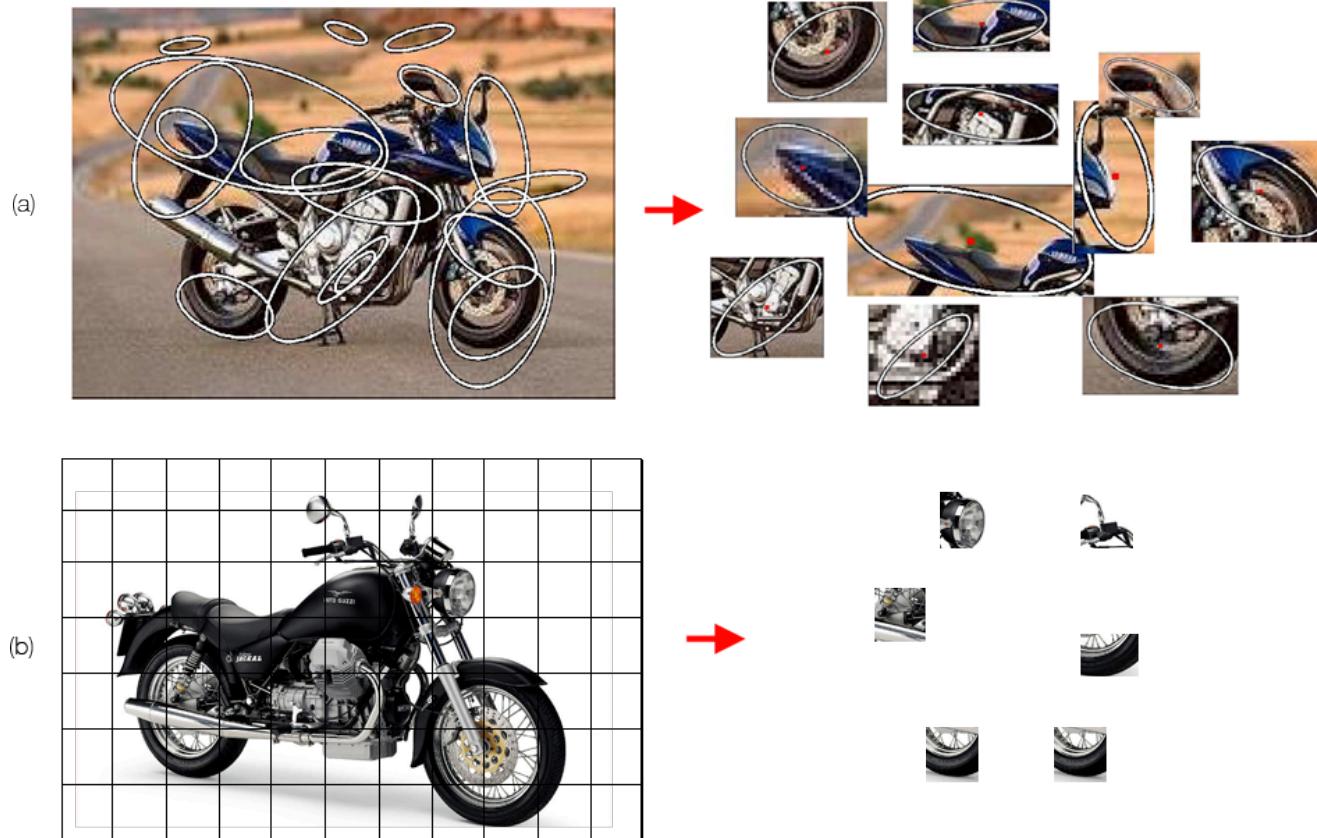


Comparativo de separação utilizando SVM e RNA. (a) Separação usando SVM (b) Separação usando RNA.  
Fonte: Produção do próprio autor.

# Conjunto de Palavras Visuais *(Bag of Visual Words) (BOVW)*

- Cada imagem é representada por um histograma com a contagem de palavras visuais existentes na imagem.
- Etapa
  1. Extração dos descritores das imagens de treinamento
    - Localização por ponto de interesse ou de forma densa.
  2. Agrupamento de descritores semelhantes em palavras visuais.
  3. Criação de histogramas associando cada descritor da imagem à palavra visual mais próxima com base em uma função de distância.
  4. Treinamento de um classificador com os histogramas.

# Conjunto de Palavras Visuais (*Bag of Visual Words*) (BOVW)



Extração de descritores de forma densa e baseada em pontos de interesse. (a) Descritores extraídos através de ponto de interesse.  
(b) Descritores extraídos de forma densa.

# Conjunto de Palavras Visuais (*Bag of Visual Words*) (BOVW)



Exemplo ilustrando a localização de descritores em diferentes imagens mas que correspondem a uma mesma característica.  
Fonte: Produção do próprio autor.

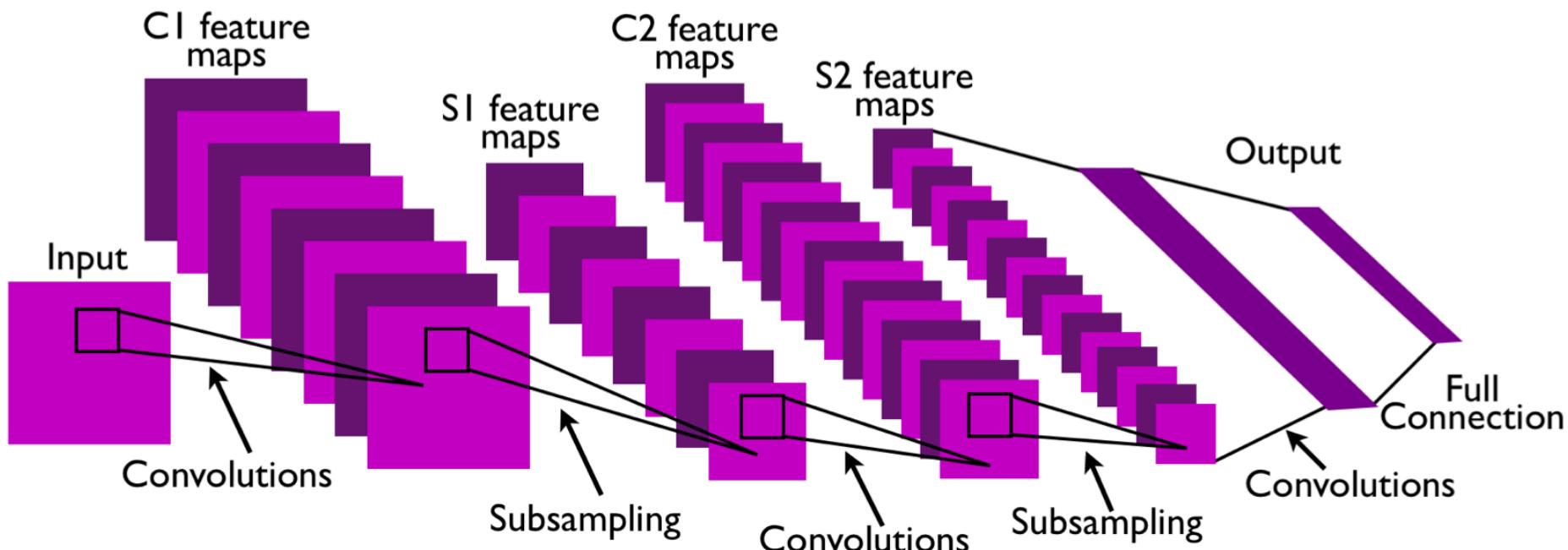
# Redes Neurais Convolucionais (CNN)

São arquiteturas biologicamente inspiradas capazes de serem treinadas e aprenderem representações invariantes (Lecum et al. 2010)

Experimentos realizados por Hubel e Wiesel (1968) sugerem uma organização hierárquica das células de tipo simples, complexas e super complexas.

- Como obter a melhor representação das características de uma imagem?
  - Com base no trabalho de Hubel, Lecum et al. sugerem uma representação hierárquica.

# Redes Neurais Convolucionais (CNN)



Rede neural convolucional com dois estágios (Lecun et al, 2010).

# Experimentos

- Conjunto de Palavras Visuais (BOVW)
- Rede Neural Convolucional (CNN)
- CNN previamente treinada como um descritor natural.

# Experimentos - BOVW

Agrupador	Classificador
<i>KMeans</i>	SVM
<i>KMeans</i>	OPF-S
OPF-U	SVM
OPF-U	OPF-S

Agrupadores e classificadores analisados.

Fonte: Produção do próprio autor.

# Experimentos - BOVW

Extrator de pontos de interesse	Descriptor
SIFT	SIFT
MSER	SIFT
Denso	SIFT
SURF	SURF
Denso	SURF

Localizadores e descritores analisados.  
Fonte: Produção do próprio autor.

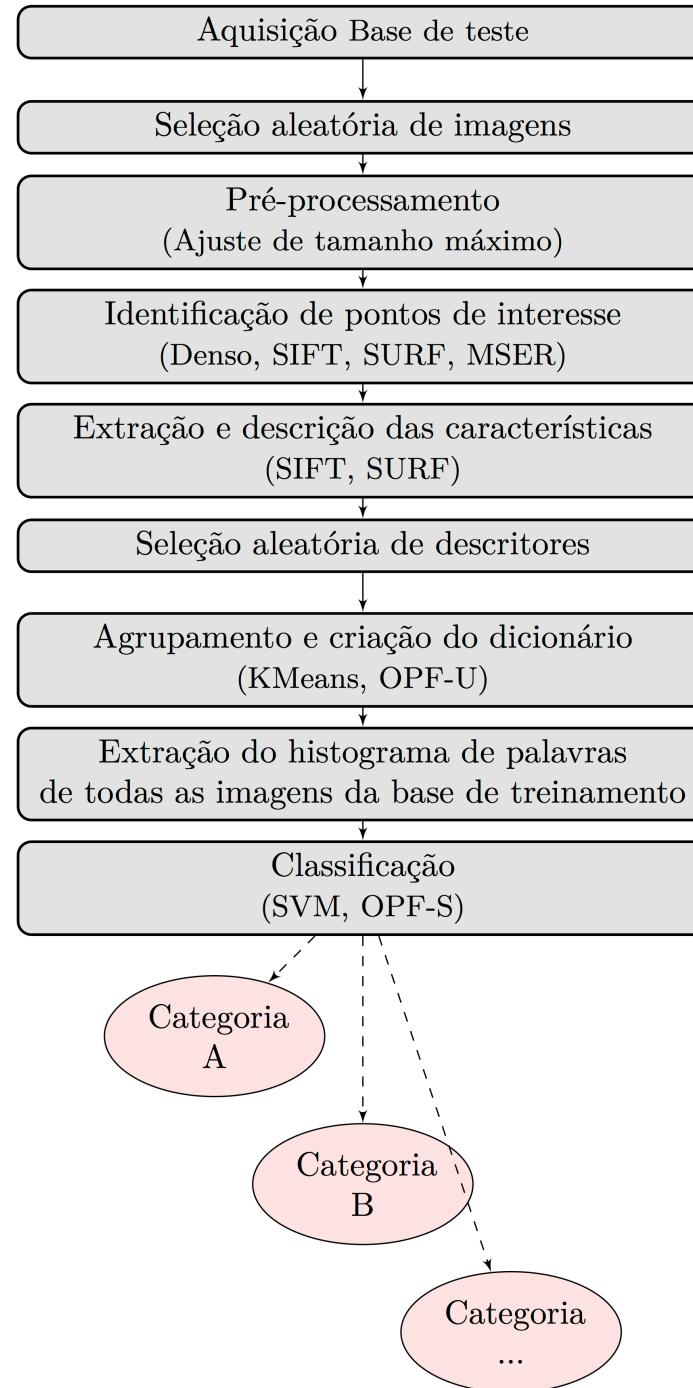
# Experimentos - BOVW

- Experimentos realizados com imagens de tamanho 64x64, 128x128, 256x256 e 512x512.
- Obtenção do melhor tamanho do dicionário de palavras visuais (K) através do OPF-U de forma empírica.
- Execução do KMeans com o mesmo K.
- Execução do KMeans com diferentes tamanhos de K.
- Criação do dicionário com 5000 descritores selecionados aleatoriamente da extração de 500 imagens, também selecionadas aleatoriamente.

# Experimentos

## BOVW

### Algoritmo



Modelo do sistema de reconhecimento  
Utilizando *KMeans* e OPF-U para agrupamento  
e SVM e OPF-S para classificação.

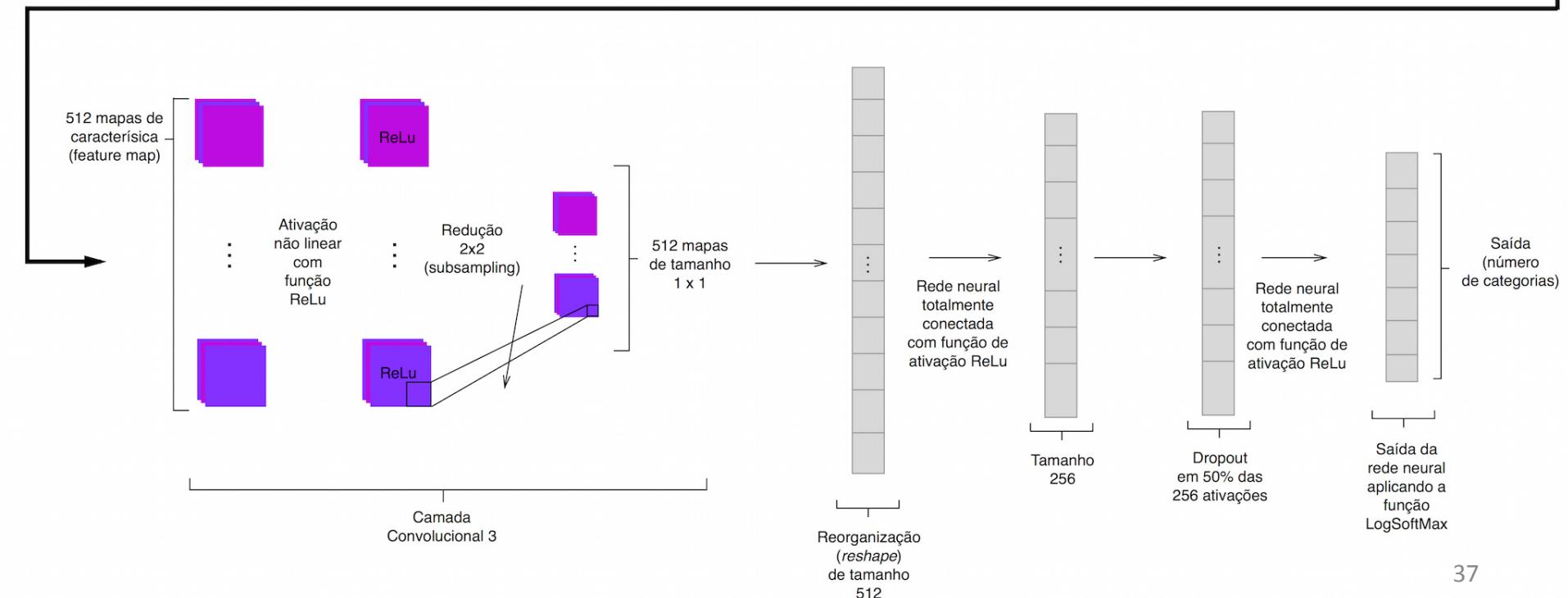
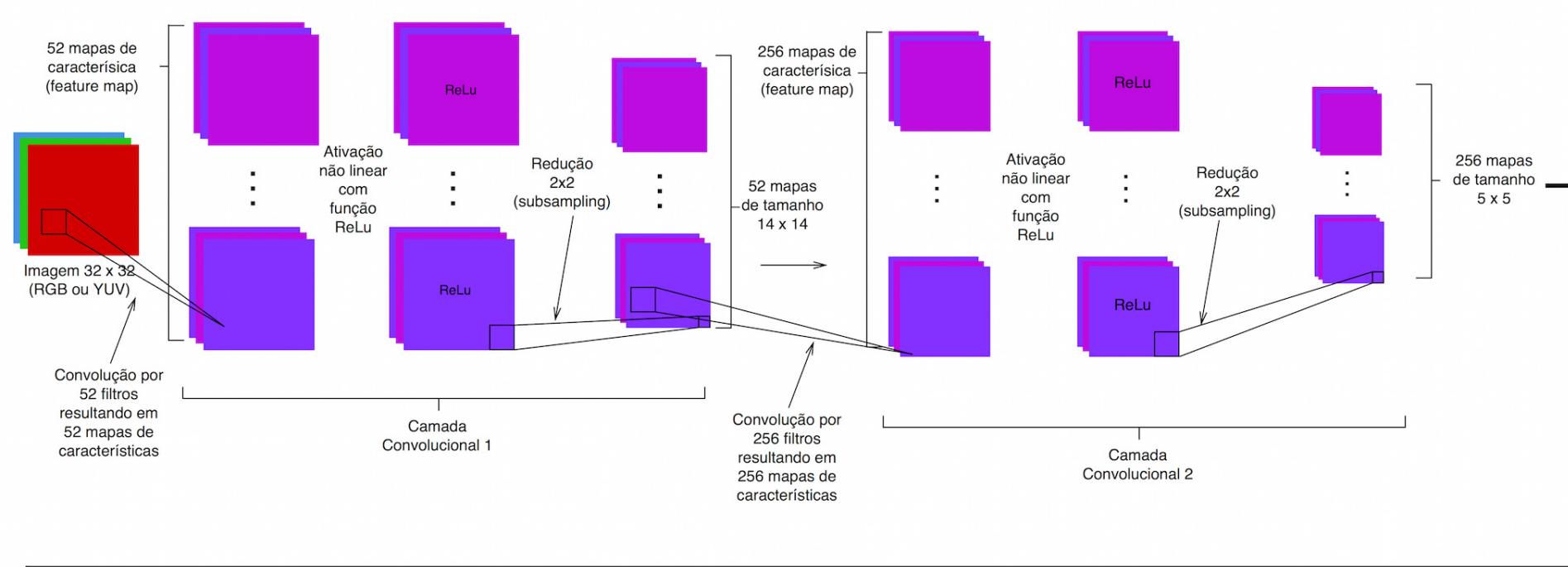
Fonte: Produção do próprio autor.

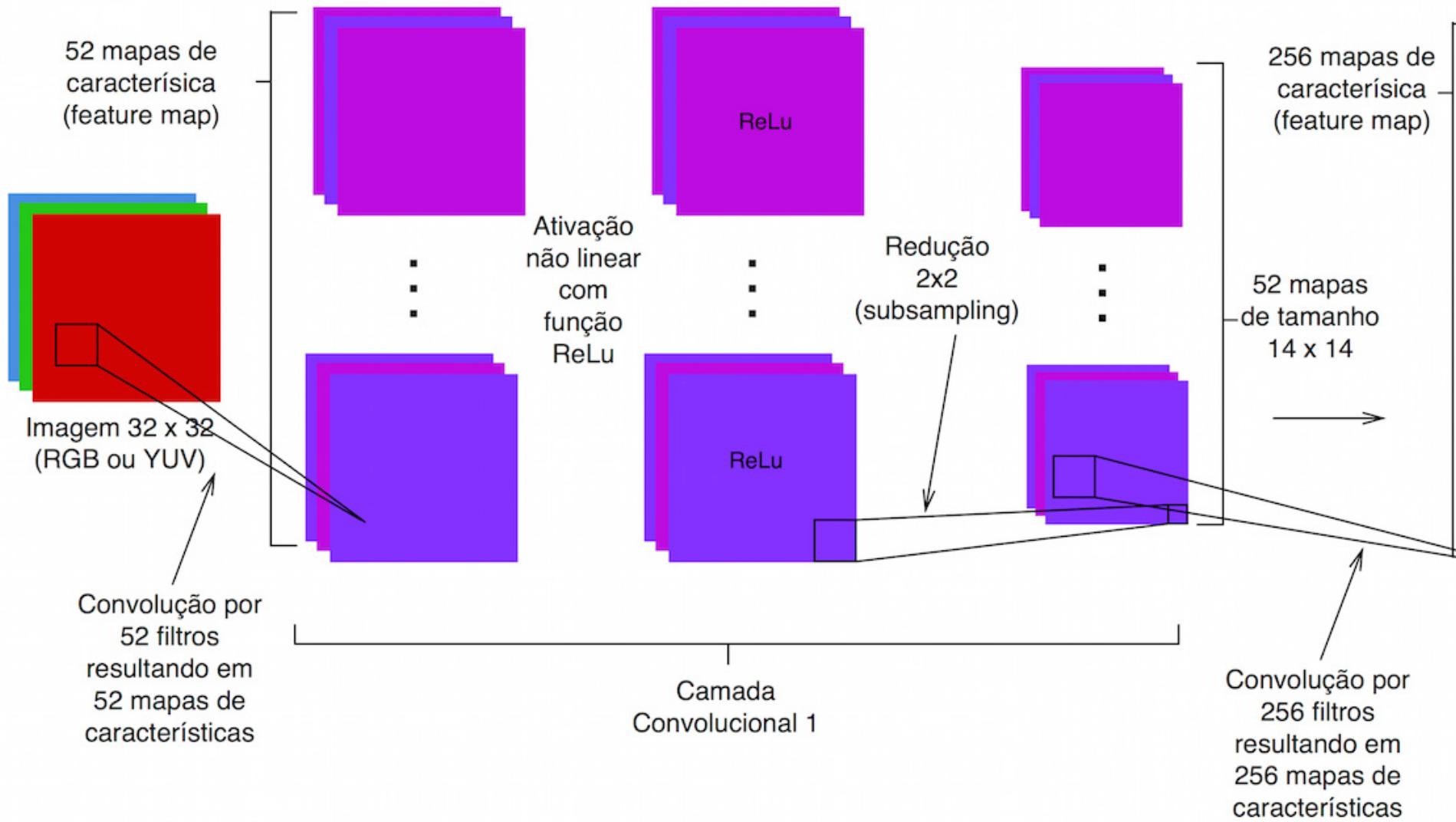
# Experimentos - CNN

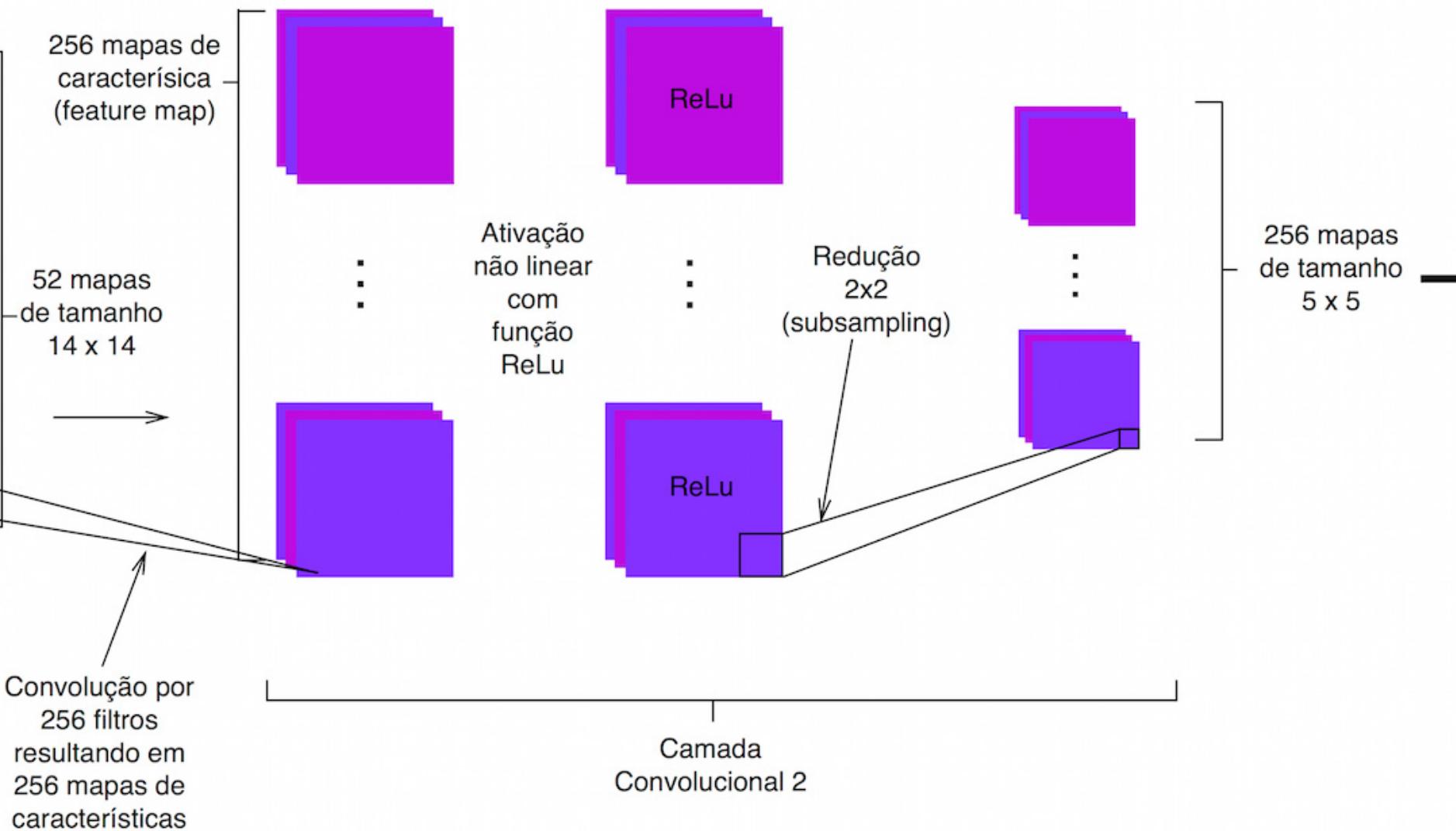
- Experimentos realizados com imagens do formato RGB e YUV.

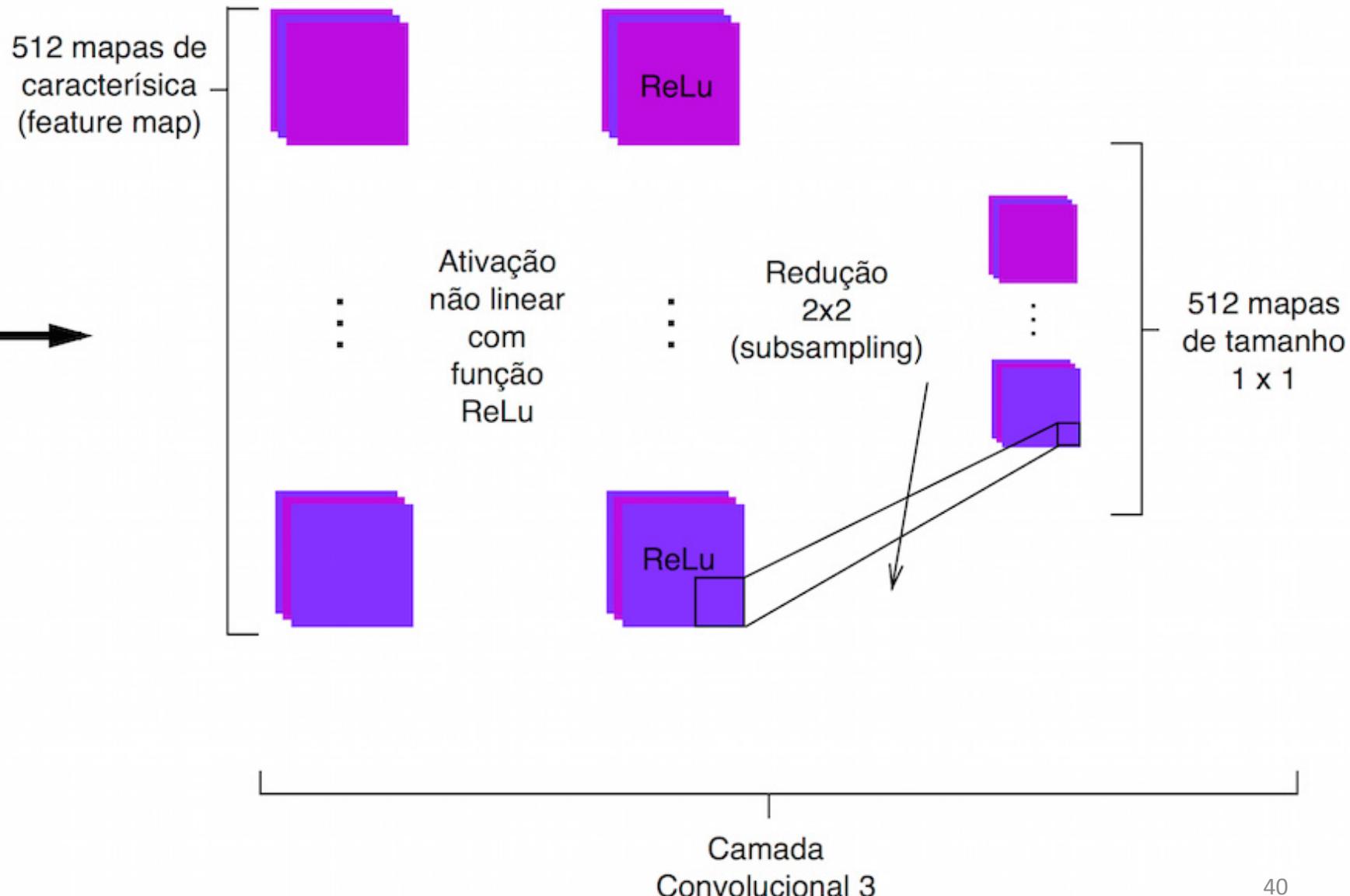
Camada	1	2	3	4	5
Tipo de estágio	Conv	Conv	Conv	Total	Total
Tam. de entrada	3	52	256	512	256
Num. de canais	52	256	512	256	no categorias
Tam. do filtro	5x5	5x5	3x3	-	-
Tam. de redução	2x2	2x2	2x2	-	-

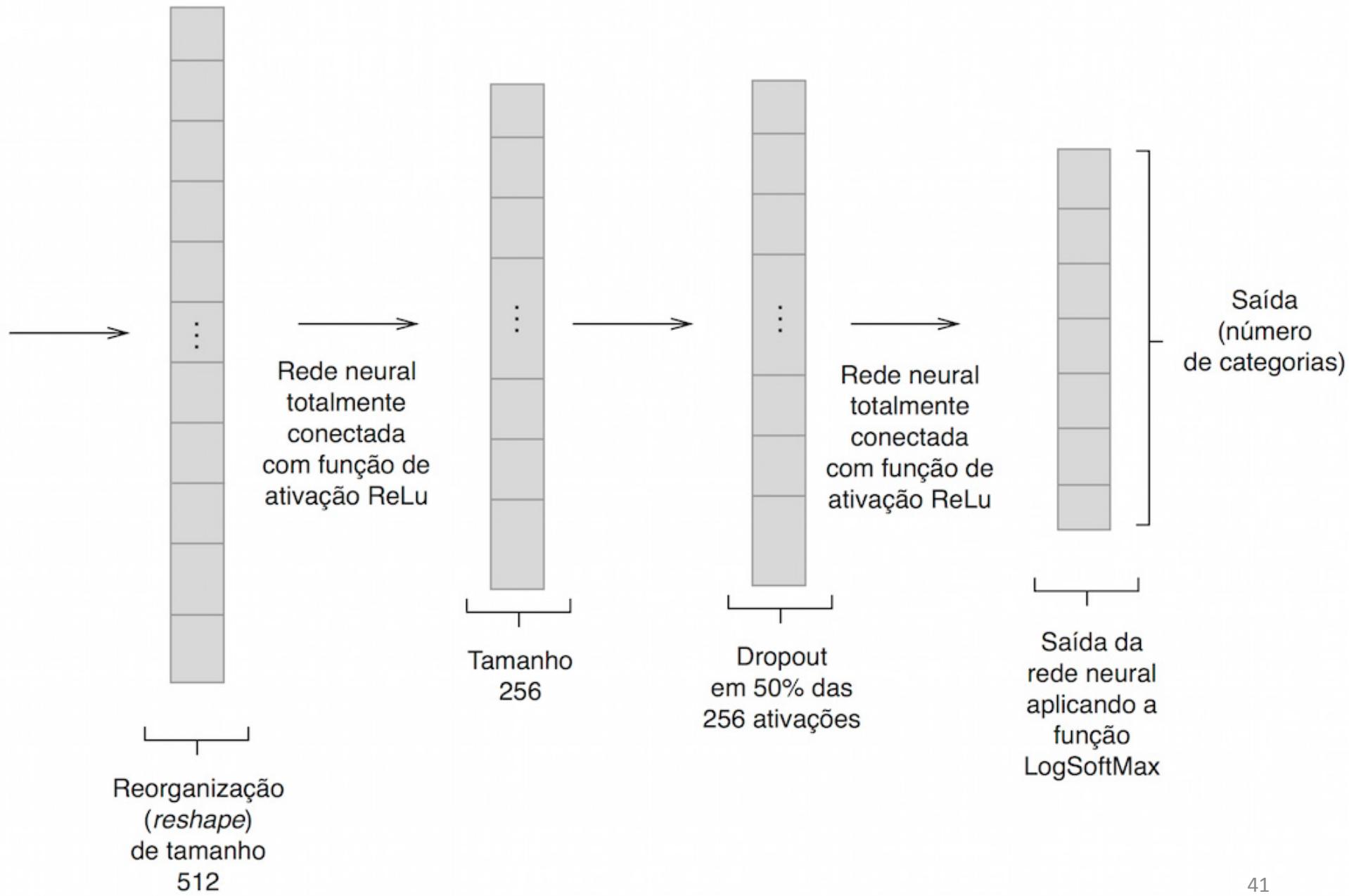
Arquitetura da CNN proposta.  
Fonte: Produção do próprio autor.



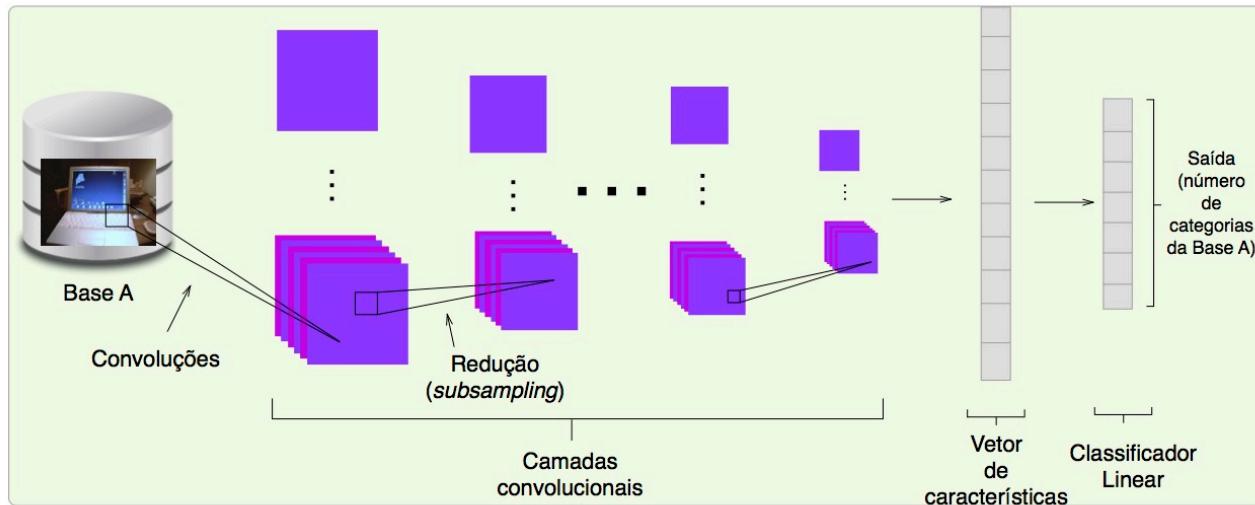






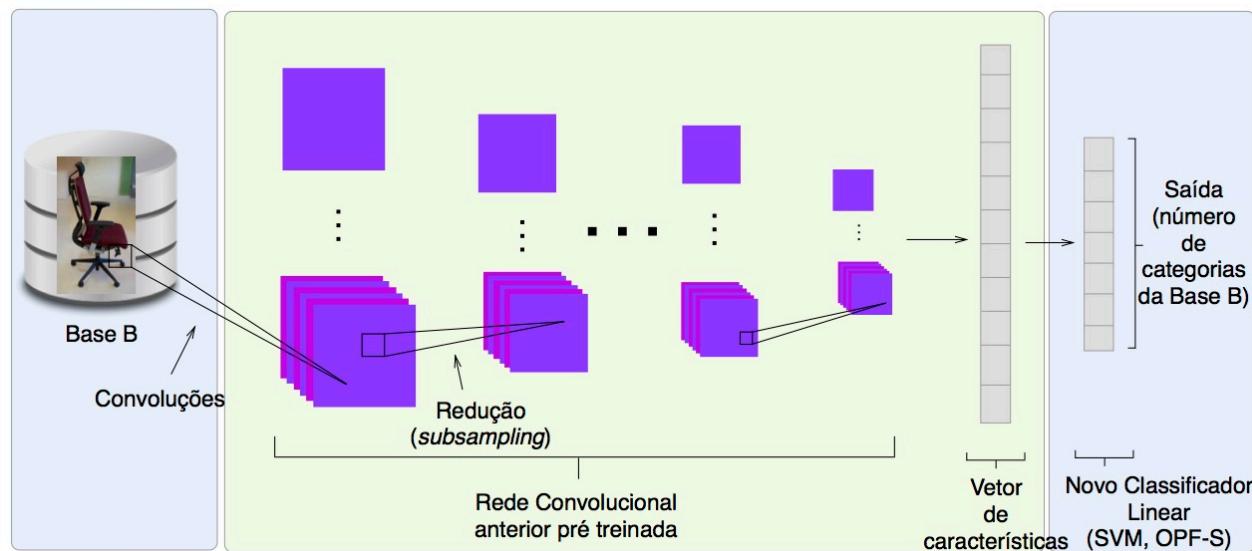
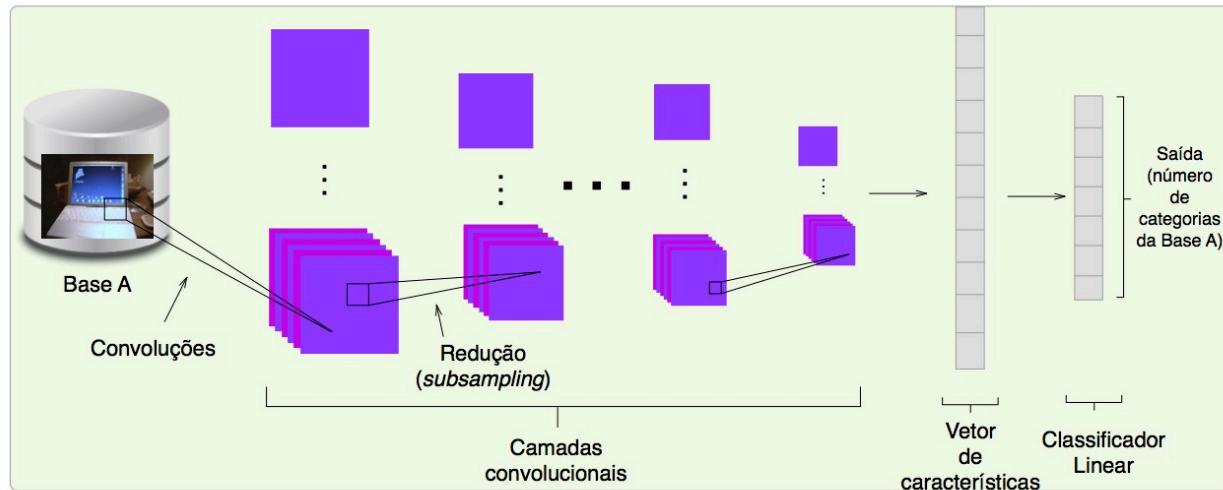


# Experimentos – CNN como descriptor natural



Camada	1	2	3	4	5	6	7	8
Tipo de estágio	Conv+Max	Conv+Max	Conv	Conv	Conv+Max	Total	Total	Total
Tam. de entrada	231x231	24x24	12x12	12x12	12x12	6x6	1x1	1x1
Num. de canais	96	256	512	1024	1024	3072	4096	1000
Tam. do filtro	11x11	5x5	3x3	3x3	3x3	-	-	-
Tam. de redução	2x2	2x2	-	-	2x2	-	-	-

# Experimentos – CNN como descriptor natural



# Base de Imagens – Caltech101

- 101 classes.
- Média de 50 imagens por classe.
- Imagens não segmentadas



Exemplos de imagens de 3 classes da base Caltech101.  
Fonte: Produção do próprio autor.

# Base de Imagens – Produtos

- Imagens de produtos extraídos de sites de comércio eletrônico.
- 36 classes.
- 350 imagens por classe.



Exemplos de imagens de 3 classes da base de imagens  
de produtos criada pelo autor.

Fonte: Produção do próprio autor.

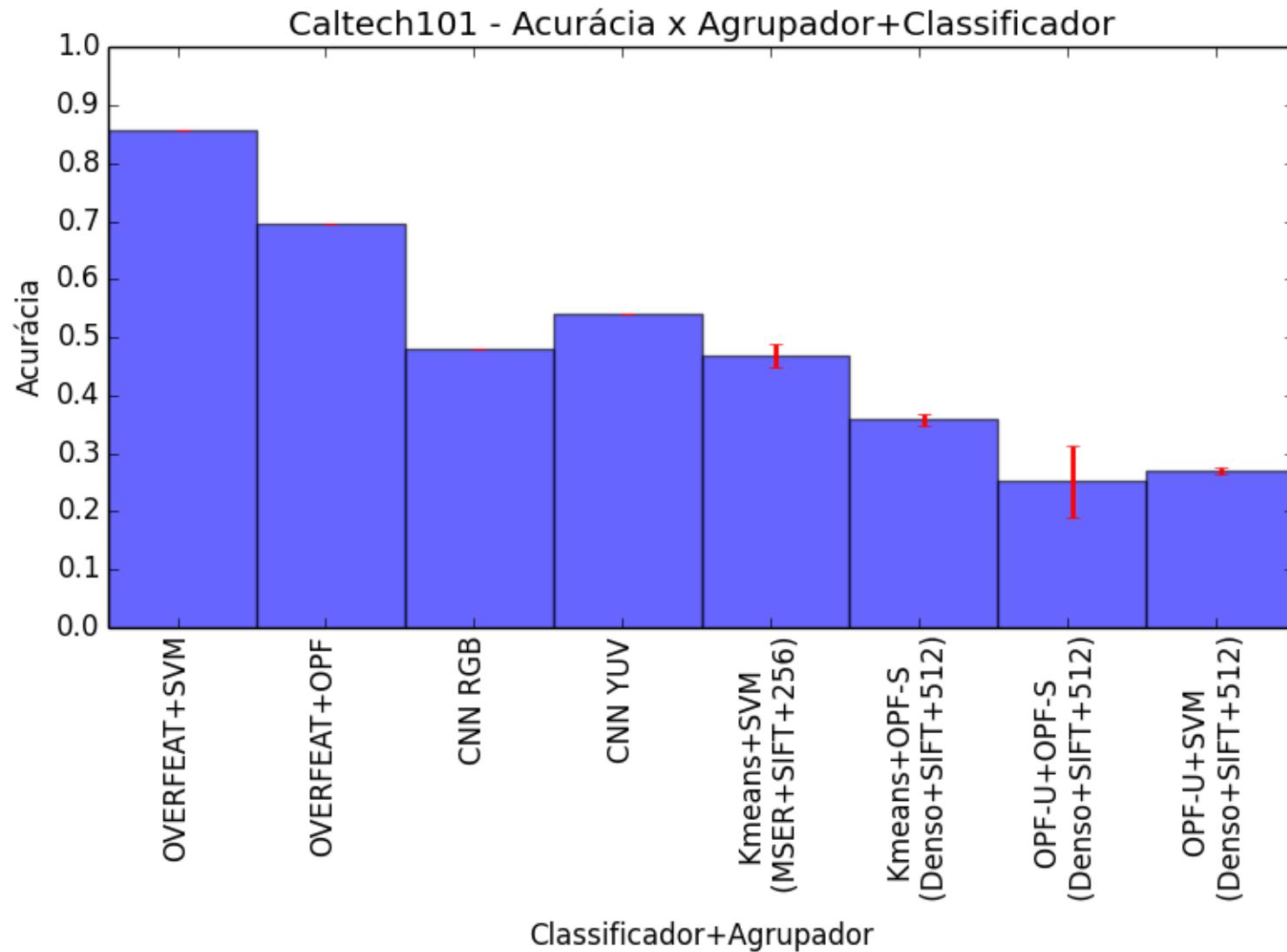
# Implementação

- OpenCV
- Torch7
- Overfeat
- Scikit Learn
- OPFLib

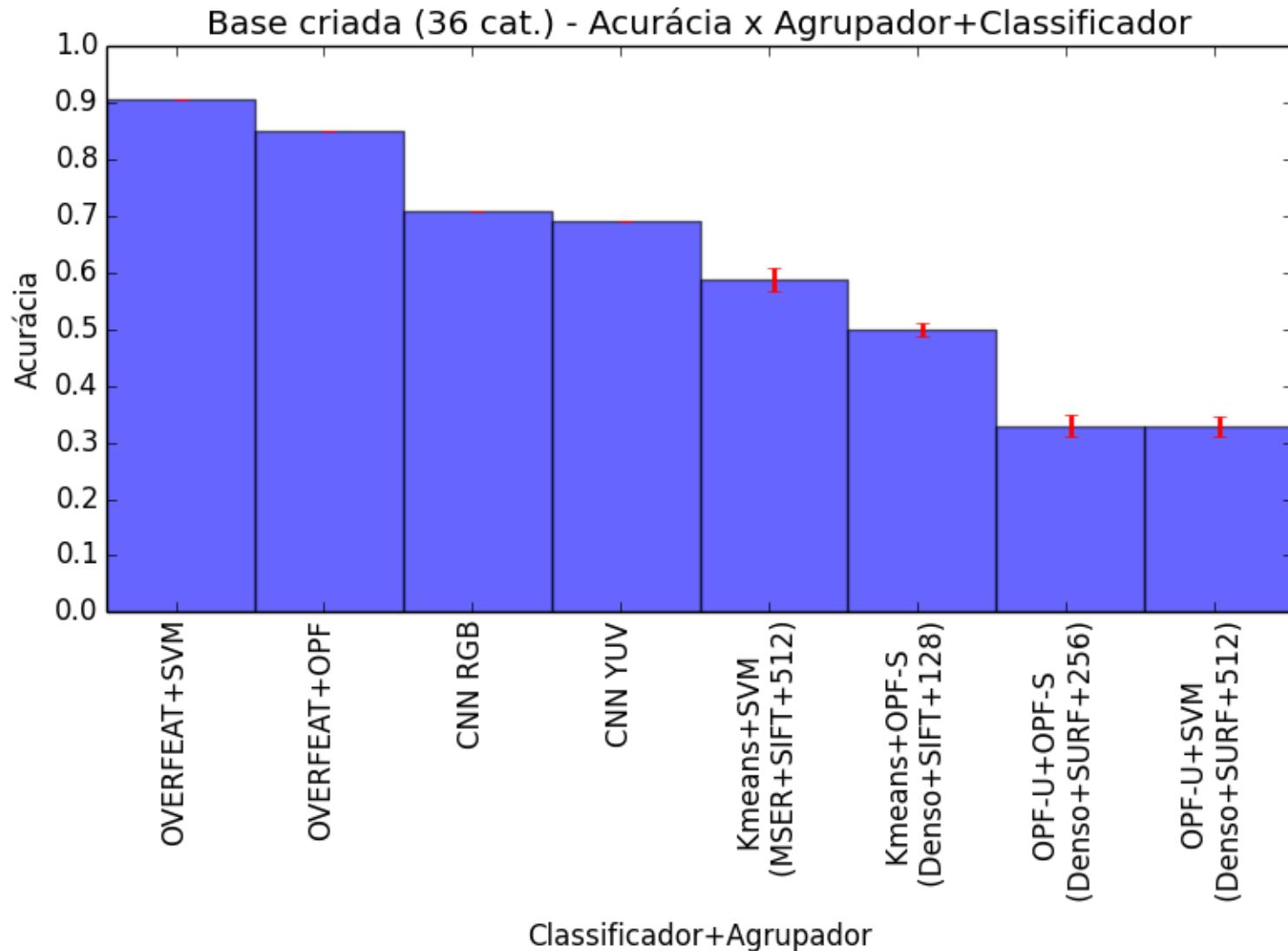
Os experimentos foram realizados no laboratório da UDESC utilizando 5 computadores idênticos em configuração e hardware e software: processador *AMD Phenom II X4 B93 2.8GHz*, 2MB L2 *cache*, com 4GB de memória DDR3 1333 MHz .

# Resultados

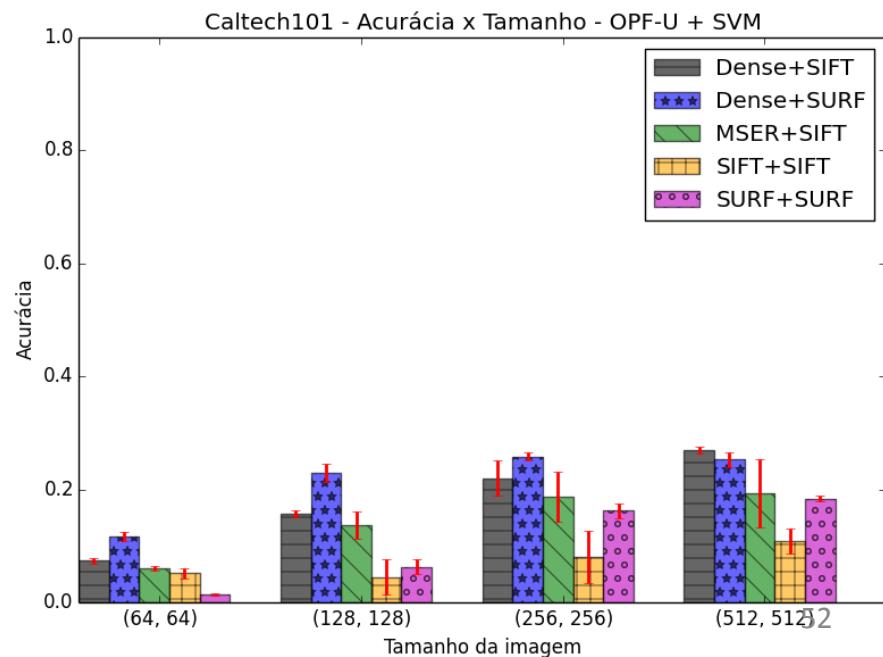
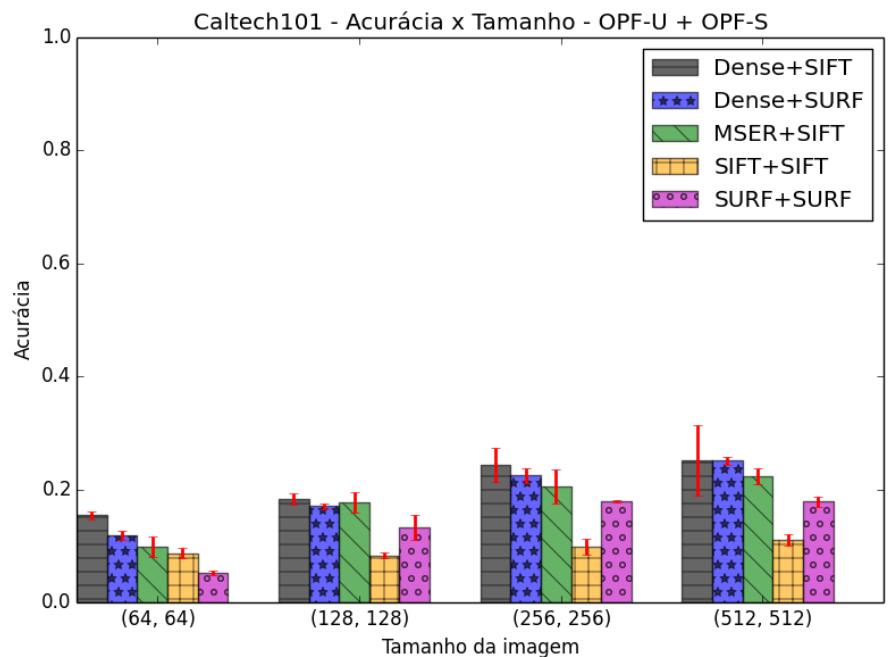
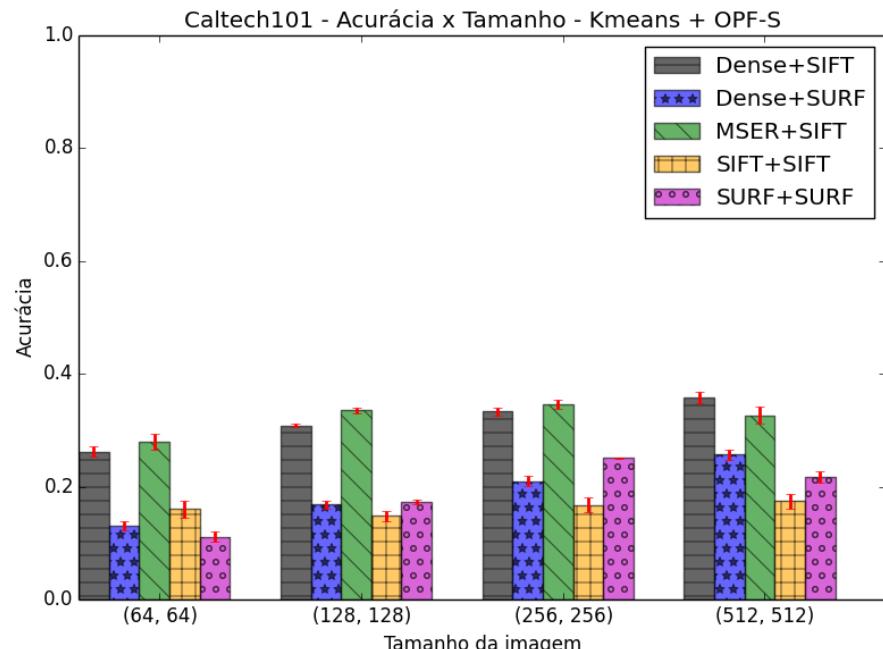
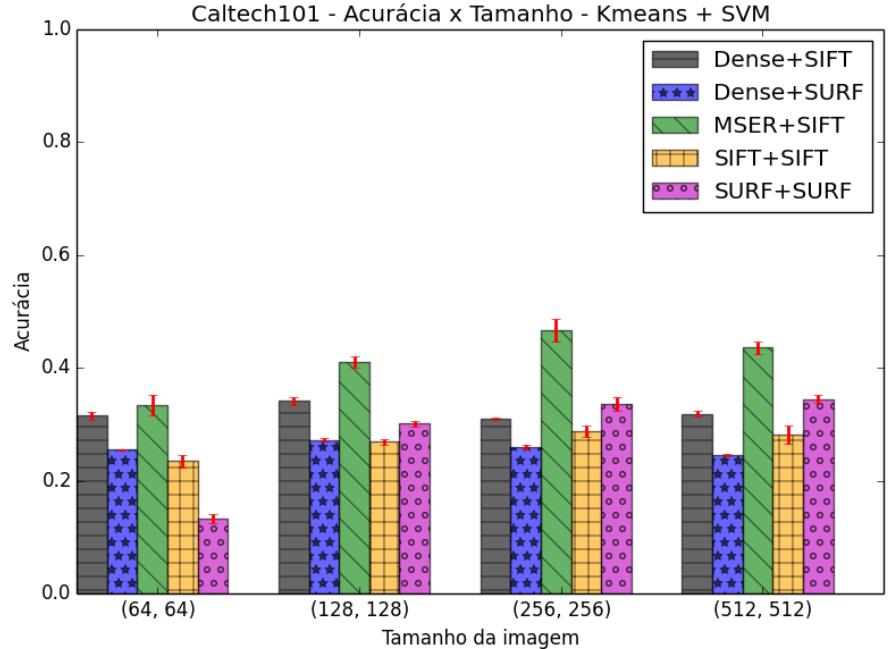
# Melhor acurácia na base Caltech101



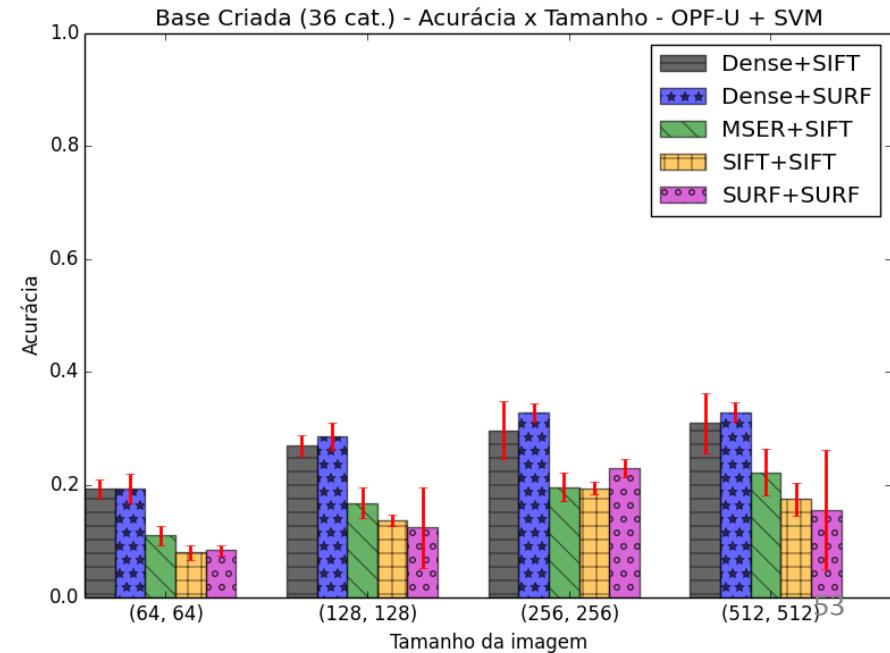
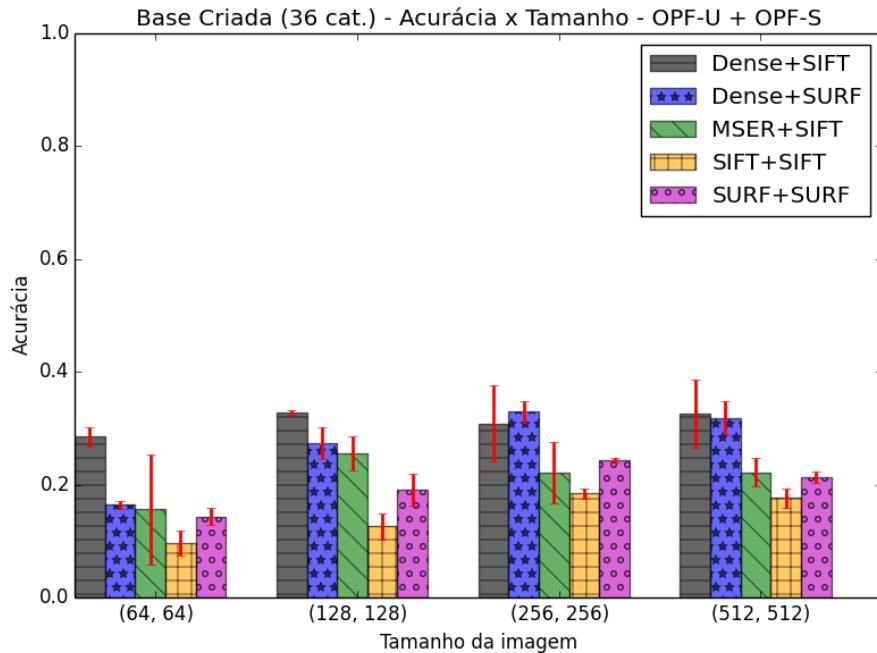
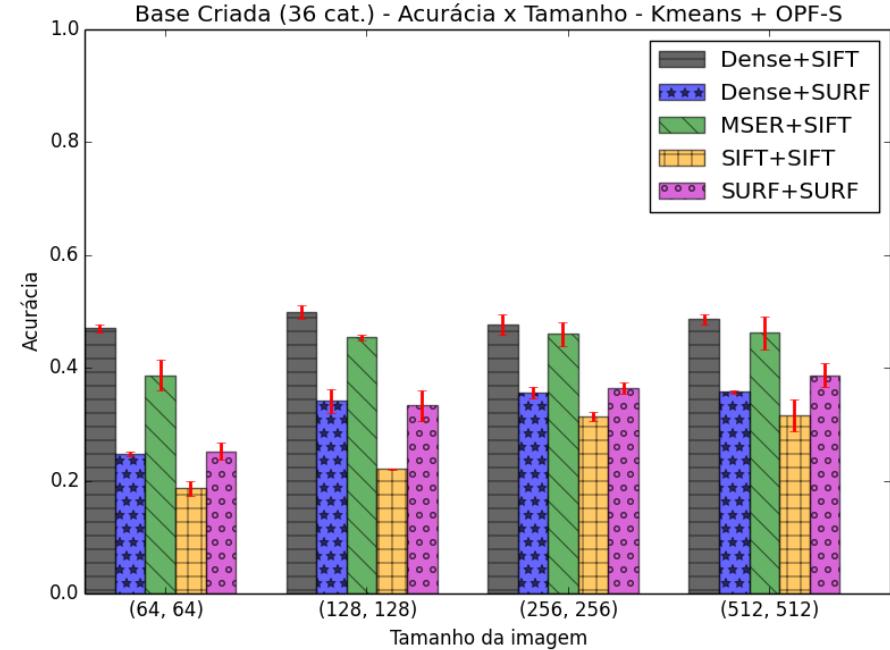
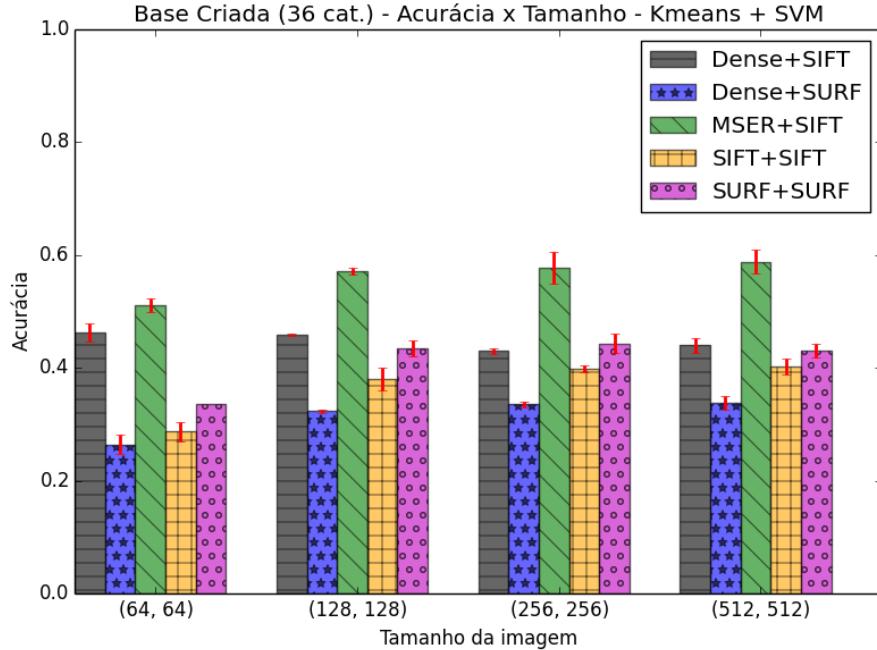
# Melhor acurácia na base de produtos



# Melhor descriptor artificial na técnica BOVW – Caltech101.



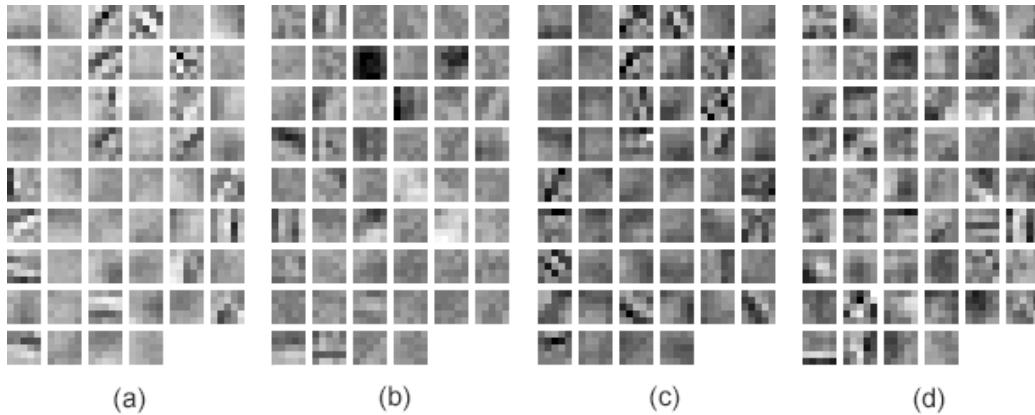
# Melhor descriptor artificial na técnica BOVW – Base de Produtos.



# Filtros da CNN



(e)



(a)

(b)

(c)

(d)

Filtros obtidos após o treinamento da rede neural. (a) Filtros da CNN YUV na base Caltech101.

(b) Filtros da CNN RGB na base Caltech101.

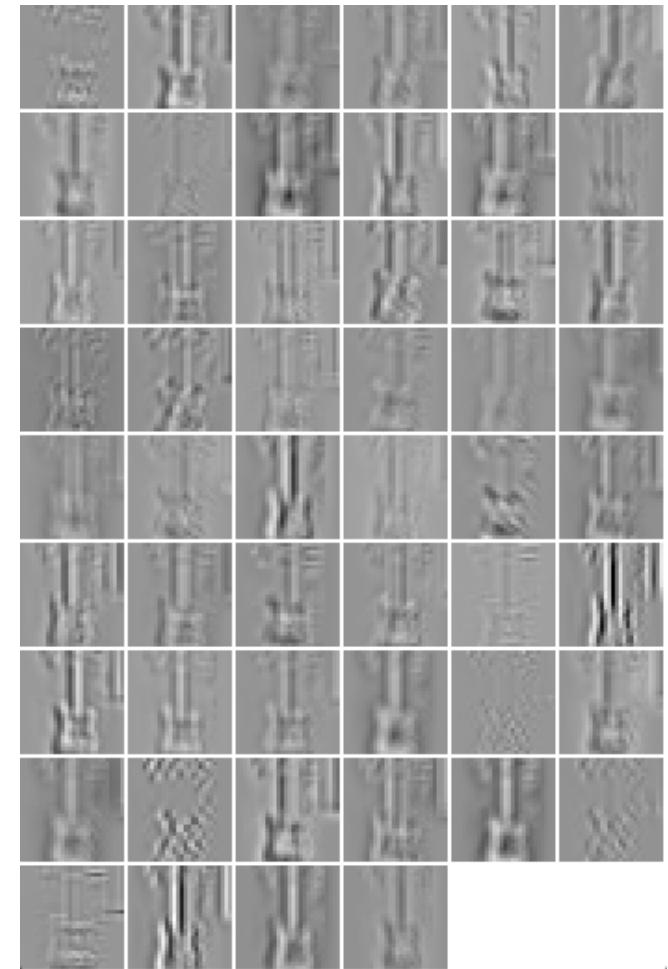
(c) Filtros da CNN YUV na base criada.

(d) Filtros da CNN RGB na base criada.

(e) Imagem de exemplo

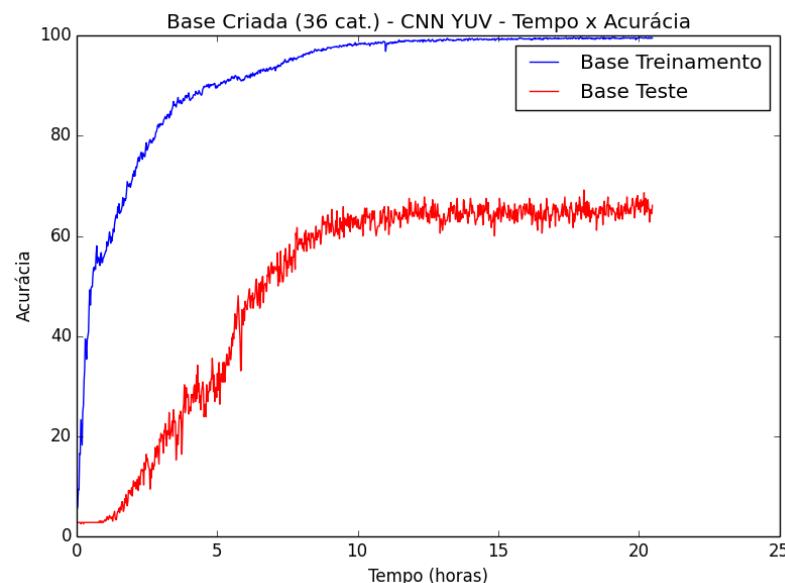
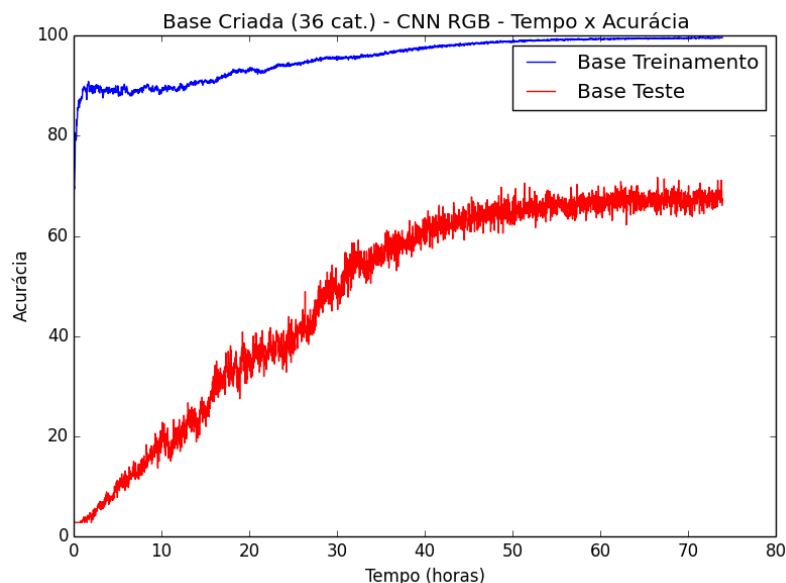
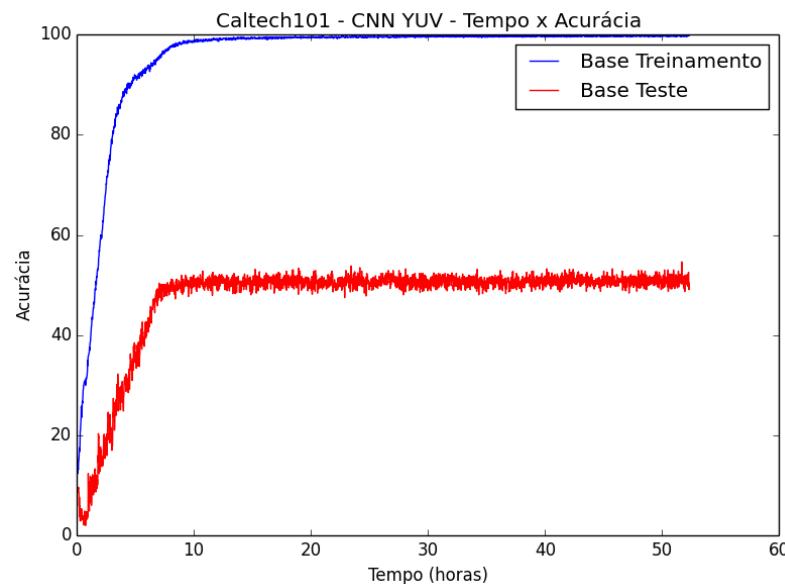
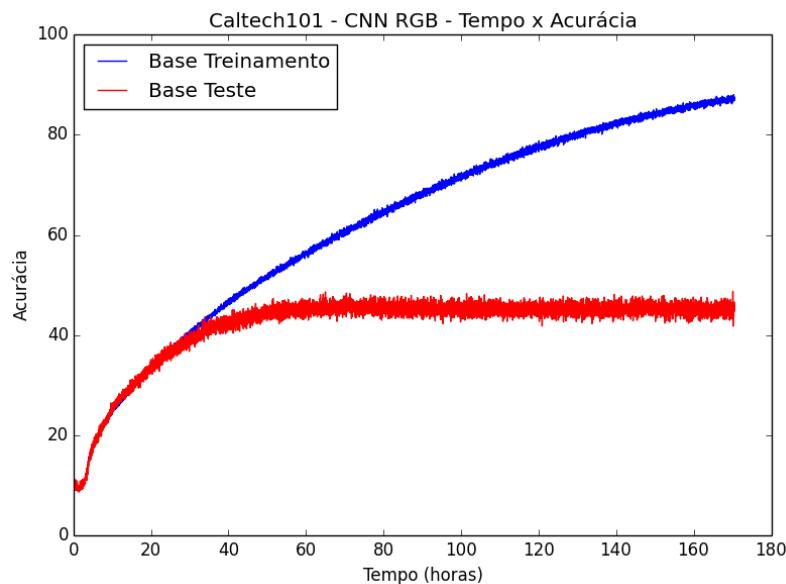
(f) Resultado da convolução dos filtros (d) na imagem (e).

Fonte: Produção do próprio autor.

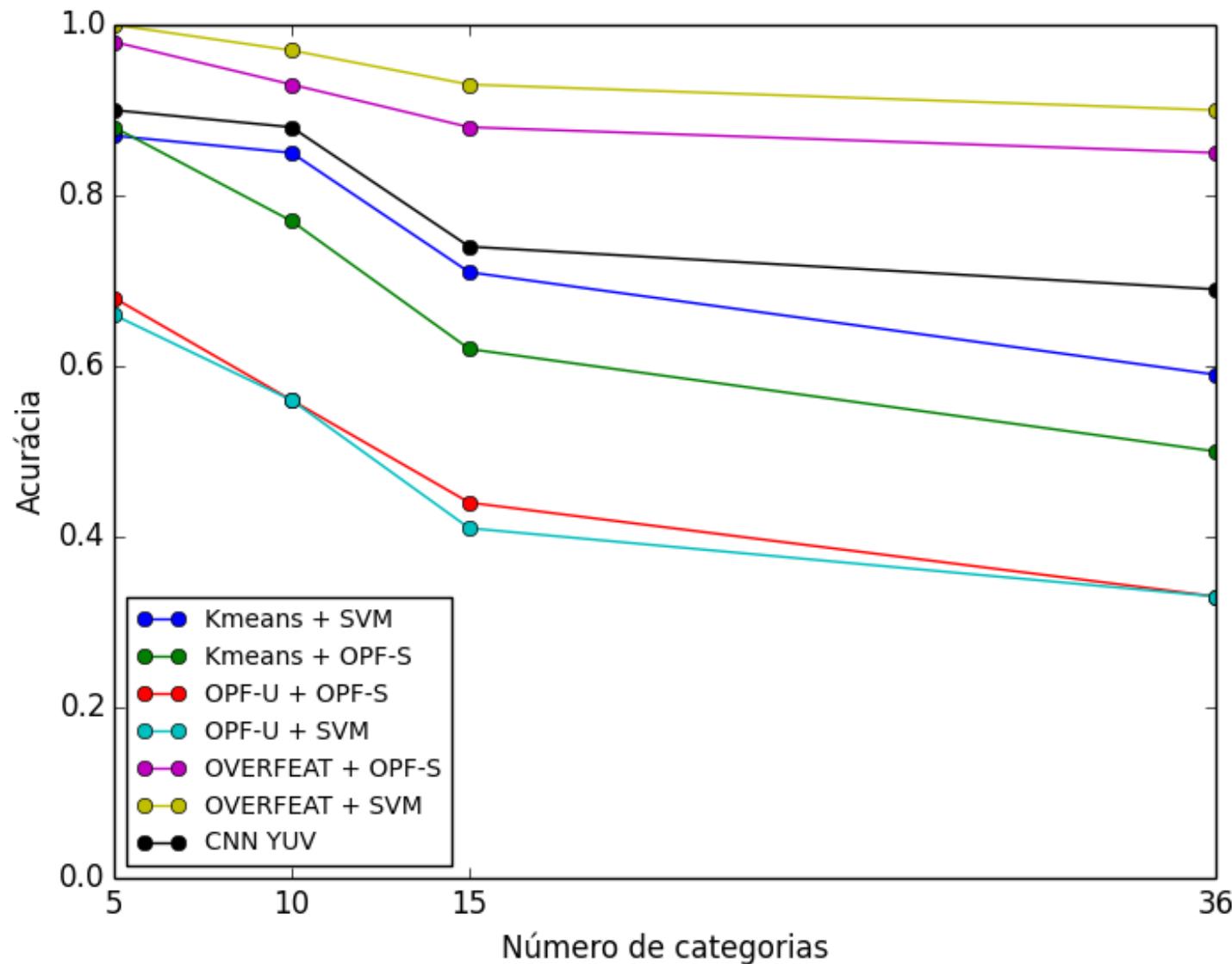


(f)

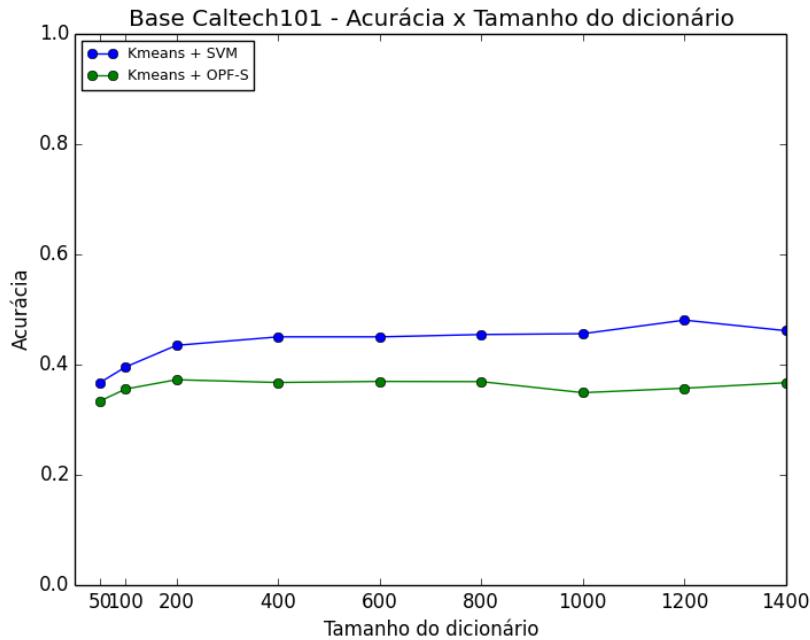
# Rede Neural Convolucional – RGB x YUV



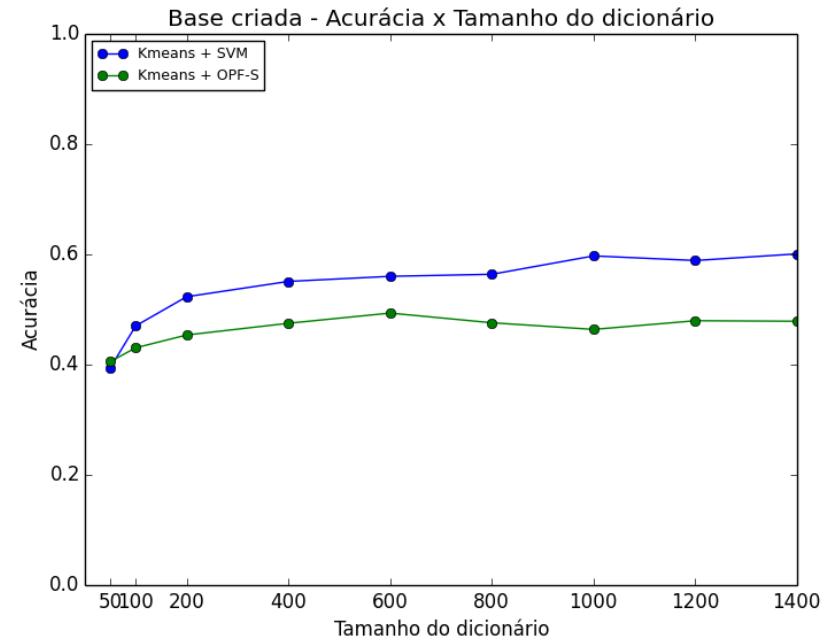
# Acurácia na base de produtos com 5, 10, 15 e 36 categorias.



# Acurácia com diferentes tamanhos de dicionário de palavras visuais (KMeans).



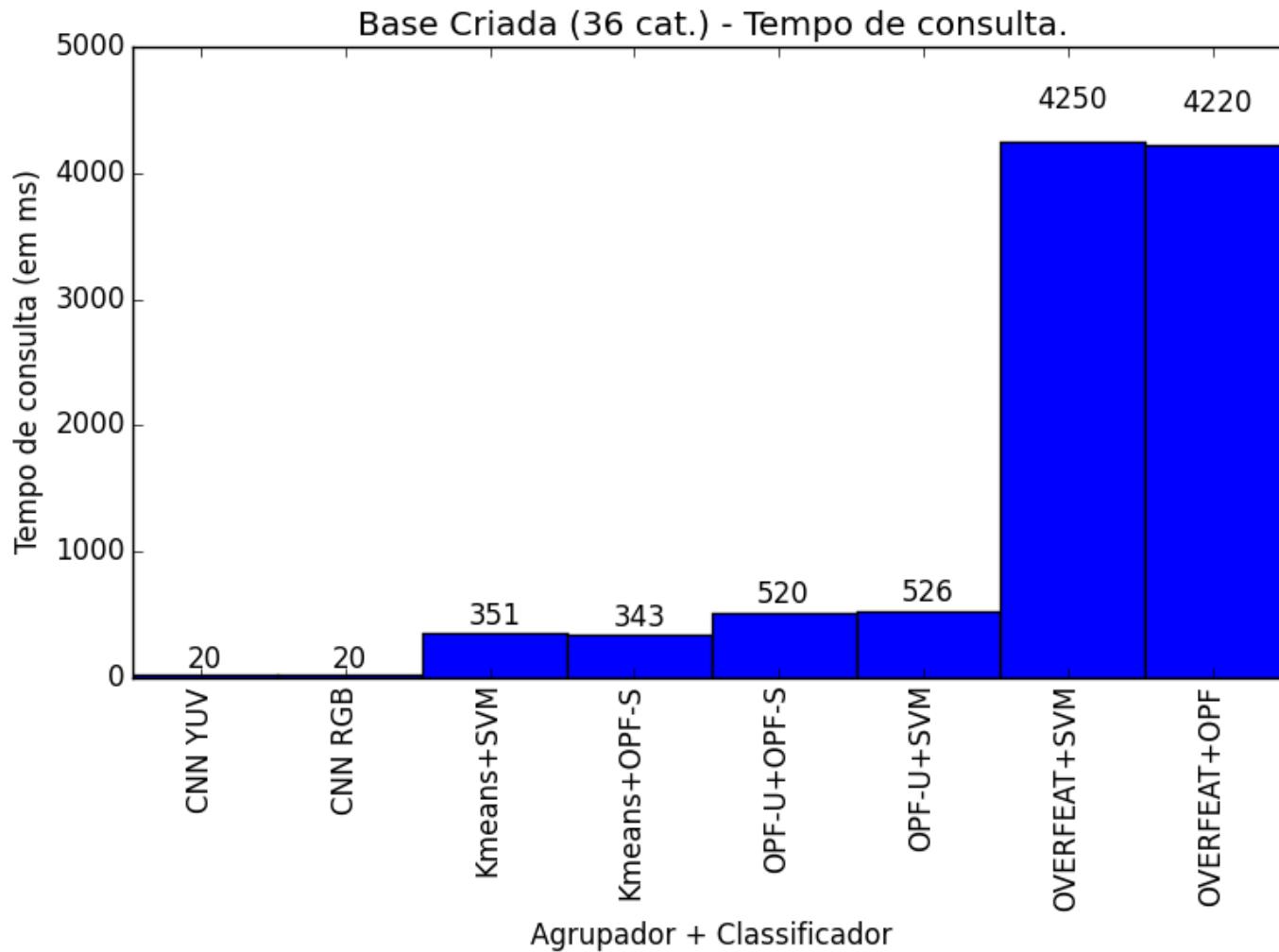
(a)



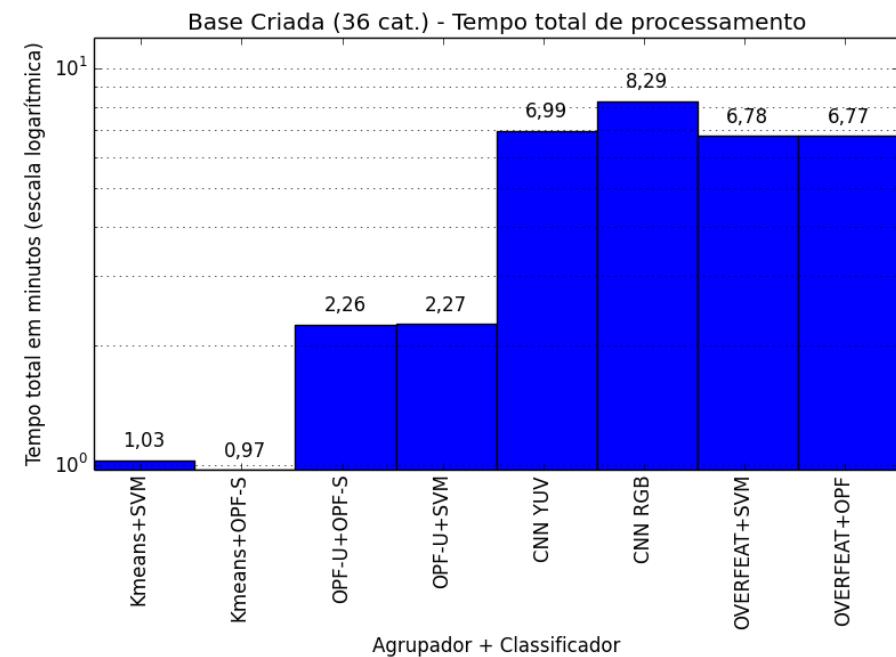
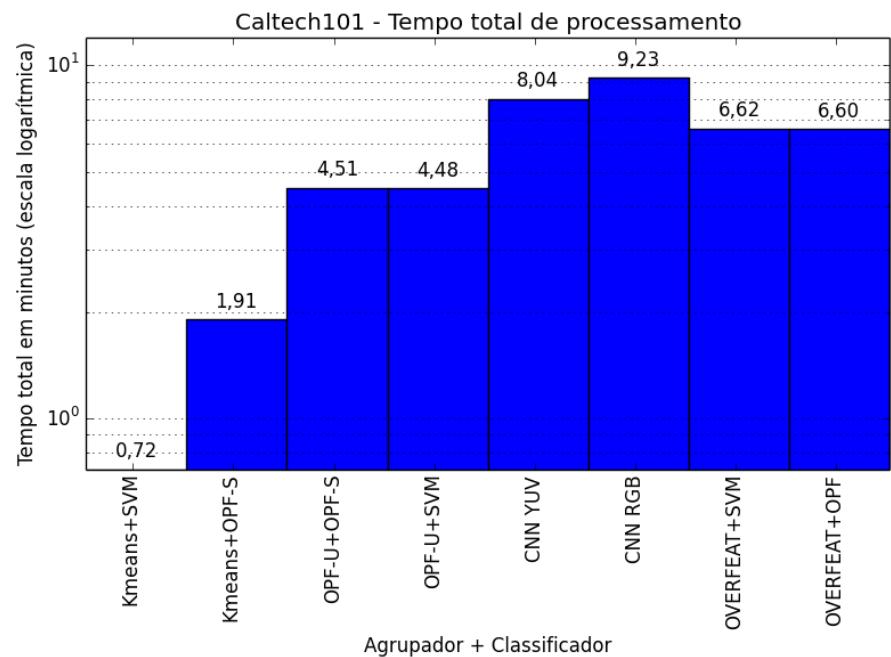
(b)

Tamanhos de dicionários no BOVW utilizando MSER+SIFT em imagens de tamanho 256: (a) resultado na base Caltech101  
(b) resultado na base de produtos  
Fonte: Produção do próprio autor.

# Tempo de classificação.



# Tempo total de processamento



Tempo total de processamento: (a) resultados na base  
Caltech101 (b) resultados na base de produtos  
Fonte: Produção do próprio autor.

# Resultados

- Melhor acurácia nas bases Caltech101 e na base de produtos.
- Melhor descriptor artificial na técnica BOVW.
- Rede Neural Convolucional – RGB x YUV
- Acurácia na base de produtos com 5, 10, 15 e 36 categorias.
- Tempo de treinamento.
- Tempo de classificação.
- Tempo total de processamento.

# Contribuições

Bases de Imagens	Técnica	Configurações			
Caltech101  Base de Produtos criada pelo autor	Conjunto de palavras visuais (BOVW)	<b>Classificador</b>	<b>Descriptor</b>	<b>Localizador</b>	<b>Tamanhos de imagem</b>
		KMeans + SVM	SIFT	Denso	64x64
		KMeans + OPF-S		MSER	128x128
		OPF-U + SVM		SIFT	256x256
		OPF-U + OPF-S	SURF	Denso	512x512
	Redes Neurais Convolucionais (CNN)	<b>Tipo de CNN</b>	<b>Configurações</b>		
		CNN sem treinamento prévio	<b>Tipos de imagem</b>		
			RGB		
			YUV		
		CNN com treinamento prévio	<b>Classificador</b>		
			SVM		
			OPF-S		

# Conclusão

- CNN como descritor natural obteve a melhor acurácia
  - 0,856 – Caltech101
  - 0,905 – Base de Produtos
- CNN sem treinamento prévio criada pelo autor obteve a segunda melhor acurácia
  - 0,540 – Caltech101
  - 0,710 – Base de Produtos
  - Acurácia muito próxima entre YUV e RGB, porém utilizando YUV o tempo de treinamento foi inferior.

# Conclusão

- Na técnica BOVW
  - Na técnica BOVW foi observada uma melhor acurácia quando utilizado o agrupador KMeans em conjunto com o classificador SVM.
  - KMeans + SVM obteve melhor acurácia do que KMeans + OPF-S
  - KMeans, de forma geral, obteve melhor acurácia do que OPF-U
  - OPF-U + SVM obteve melhor acurácia do que OPF-U + OPF-S
  - O localizador MSER em conjunto com o descriptor SIFT obtiveram a melhor acurácia.
  - Algoritmos utilizando o OPF obtiveram melhor acurácia quando os descritores foram extraídos de forma densa.

# Trabalhos Futuros

- Utilizar uma CNN pré-treinada e realizar o ajuste dos filtros como um refinamento em uma nova base.
- Verificar se a transferência de conhecimento (reutilização dos filtros pré-treinados) é aplicável a problemas de classificação binários ou de nichos específicos.
  - Ex: utilização da Overfeat para classificação de placas de trânsito, imagens médicas, plantas entre outros.
- BOVW com HOG
  - Extraídos de forma densa
  - Extraídos com base em pontos de interesse (Ex: MSER)
- BOVW com combinação de descritores e classificadores.
- Comparar o BOVW com pirâmide de escalas e localização das palavras visuais.

# Referências

- AFONSO, L. C. S.; PAPA, J. P.; PAPA, L. P.; MARANA, A. N.; ROCHA, A. Automatic visual dictionary generation through optimum-path forest clustering. In: *ICIP*. IEEE, 2012. p. 1897–1900. ISBN 978-1-4673-2534-9.
- BAY, H.; ESS, A.; TUYTELAARS, T.; GOOL, L. V. Speeded- up robust features (surf). *Comput. Vis. Image Underst.*, Elsevier Science Inc., New York, NY, USA, v. 110, n. 3, p. 346–359, jun. 2008. ISSN 1077-3142.
- CIRESAN, D.; MEIER, U.; SCHMIDHUBER, J. Multi-column deep neural networks for image classification. In: IN PROCEEDINGS OF THE 25TH IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR 2012. [S.l.: s.n.], 2012. p. 3642–3649.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Mach. Learn.*, Kluwer Academic Publishers, Hingham, MA, USA, v. 20, n. 3, p. 273–297, set. 1995. ISSN 0885-6125.

# Referências

DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. [S.l.: s.n.], 2005. v. 1, p. 886–893 vol. 1. ISSN 1063-6919.

DATTA, R.; JOSHI, D.; LI, J.; WANG, J. Z. Image retrieval: ideas, influences, and trends of the new age. *ACM COMPUTING SURVEYS*, 2008.

GOODFELLOW, I. J.; WARDE-FARLEY, D.; MIRZA, M.; COURVILLE, A. C.; BENGIO, Y. Maxout networks. *CoRR*, abs/1302.4389, 2013.

HINTON, G. E.; SRIVASTAVA, N.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012.

HUBEL, D. H.; WIESEL, T. N. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, v. 195, n. 1, p. 215–243, 1968.

# Referências

JARRETT, K.; KAVUKCUOGLU, K.; LECUN, Y. What is the best multi-stage architecture for object recognition? IEEE 12th International Conference on Computer Vision, 2009.

KAVUKCUOGLU, K.; SERMANET, P.; BOUREAU, Y. Ian; GREGOR, K.; MATHIEU, M.; LECUN, Y. *Learning Convolutional Feature Hierarchies for Visual Recognition*. 2012.

LAZEBNIK, S.; SCHMID, C.; PONCE, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2. Washington, DC, USA: IEEE Computer Society, 2006. (CVPR '06), p. 2169–2178. ISBN 0-7695-2597-0.

LI, F.-F.; PERONA, P. A bayesian hierarchical model for learning natural scene categories. In: *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2 - Volume 02*. Washington, DC, USA: IEEE Computer Society, 2005. (CVPR '05), p. 524–531. ISBN 0-7695-2372-2.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. Convolutional networks and applications in vision. In: *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*. [S.l.: s.n.], 2010. p. 253–256.

# Referências

LOWE, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, Kluwer Academic Publishers, Hingham, MA, USA, v. 60, n. 2, p. 91–110, nov. 2004. ISSN 0920-5691.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. Berkeley, Calif.: University of California Press, 1967. p. 281–297.

MATAS, J.; CHUM, O.; URBAN, M.; PAJDLA, T. Robust wide baseline stereo from maximally stable extremal regions. In: *Proceedings of the British Machine Vision Conference*. [S.l.]: BMVA Press, 2002. p. 36.1–36.10. ISBN 1-901725-19-7. Doi:10.5244/C.16.36.

MIKOŁAJCZYK, K.; TUYTELAARS, T.; SCHMID, C.; ZISSERMAN, A.; MATAS, J.; SCHAFFALITZKY, F.; KADIR, T.; GOOL, L. V. A comparison of ane region detectors. *Int. J. Comput. Vision*, Kluwer Academic Publishers, Hingham, MA, USA, v. 65, n. 1-2, p. 43–72, nov. 2005. ISSN 0920-5691.

TANG, Y. Deep learning using support vector machines. *CoRR*, abs/1306.0239, 2013.

# Referências

PAPA, J. P.; FALCÃO, A. X.; SUZUKI, C. T. N. Supervised pattern classification based on optimum-path forest. *Int. J. Imaging Syst. Technol.*, John Wiley & Sons, Inc., New York, NY, USA, v. 19, n. 2, p. 120–131, jun. 2009. ISSN 0899-9457.

RAZAVIAN, A. S.; AZIZPOUR, H.; SULLIVAN, J.; CARLSSON, S. Cnn features o-the-shelf: an astounding baseline for recognition. *CoRR*, abs/1403.6382, 2014.

ROSTEN, E.; DRUMMOND, T. Machine learning for high-speed corner detection. In: *In European Conference on Computer Vision*. [S.I.: s.n.], 2006. p. 430–443.

RUBLEE, E.; RABAUD, V.; KONOLIGE, K.; BRADSKI, G. Orb: An efficient alternative to sift or surf. In: *Proceedings of the 2011 International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 2011. (ICCV '11), p. 2564–2571. ISBN 978-1-4577-1101-5.

# Referências

SERMANET, P.; EIGEN, D.; ZHANG, X.; MATHIEU, M.; FERGUS, R.; LECUN, Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. *CoRR*, abs/1312.6229, 2013.

TORRES, R. D. S.; FALCÃO, A. X. Content-based image retrieval: Theory and applications. *Revista de Informática Teórica e Aplicada*, v. 13, p. 161–185, 2006.

WALLRAVEN, C.; CAPUTO, B.; GRAF, A. Recognition with local features: the kernel recipe. In: *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*. Washington, DC, USA: IEEE Computer Society, 2003. (ICCV '03), p. 257–. ISBN 0-7695-1950-4.

ZEILER, M. D.; FERGUS, R. Stochastic pooling for regularization of deep convolutional neural networks. *CoRR*, abs/1301.3557, 2013.

# Obrigado!

