

1.01.04 - Matemática / Matemática Aplicada

CLASSIFICAÇÃO DE IMAGENS BIOLÓGICAS USANDO DESCRITORES EXTRAÍDOS POR REDES CONVOLUCIONAIS PROFUNDAS

Marina Rocha Martins Monteiro^{1*}, João Batista Florindo²

1. Estudante do Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas (IMECC-UNICAMP)
2. Professor Doutor do IMECC-UNICAMP – Departamento de Matemática Aplicada/Orientador

Resumo

Redes neurais convolucionais têm sido muito relevantes na área de classificação. Neste projeto, elas foram utilizadas na categorização de imagens biológicas, por meio da extração de descritores e votação entre diferentes classificadores (*ensemble*), utilizando estratégias de combinação de descritores de baixo e alto nível para aumentar a acurácia. A metodologia desenvolvida foi testada na base “KTH-TIPS2-2b” [10], confirmando sua eficiência diante de problemas típicos de classificação de imagens, então ela foi aplicada a dois problemas de grande importância. O primeiro é a identificação de espécies de plantas brasileiras a partir de fotos da superfície foliar [2]. O segundo é a categorização de tipos de cistos bucais [4]. O *ensemble* resultou em um aumento na taxa de acerto e a estratégia de combinação de descritores se mostrou promissora, resultando em um sistema computacional capaz de auxiliar no trabalho de médicos e botânicos na categorização de cistos bucais e espécies de plantas.

Palavras-chave: redes neurais; *ensemble* de classificadores; combinação de *features*.

Apoio financeiro: PIBIC - CNPq.

Trabalho selecionado para a JNIC: PRP - Pró-Reitoria de Pesquisa da Unicamp.

Introdução

As redes neurais são uma poderosa ferramenta computacional, que permeia praticamente todos os aspectos da sociedade moderna: desde buscadores *online* até filtros de conteúdos em redes sociais e muitos outros exemplos. Atualmente, o interesse nessas redes vem crescendo significativamente, principalmente nas redes convolucionais profundas. Por ser um método que vem mostrando desempenho superior a outros modelos de classificação, tem sido aplicado em diversas áreas. Algumas dessas aplicações são, por exemplo, em imagens médicas em geral [9], na predição de sequências de DNA-RNA [1], entre muitas outras.

Uma característica importante das redes neurais é que elas são fortemente hierarquizadas e flexíveis, o que as torna capazes de extrair descritores de imagens mesmo que estas sejam bastante complexas [6]. Este projeto utilizou essa propriedade na extração de descritores em redes convolucionais profundas e os combinou com diversos classificadores diferentes, além de ter aplicado uma nova técnica de combinação de descritores de diferentes níveis para aumentar a taxa de acerto.

A metodologia desenvolvida foi aplicada a dois problemas práticos de grande relevância e que envolvem a análise de imagens biológicas. O primeiro é a identificação de espécies de plantas brasileiras a partir de imagens escaneadas da superfície foliar [2]. Este é um problema naturalmente desafiador, uma vez que as estruturas classicamente usadas pelos especialistas para este fim são flores e frutos, que muitas vezes não são facilmente acessíveis na natureza, portanto a possibilidade de se usar a folha é um grande facilitador para este processo. O segundo problema é a identificação e categorização de três tipos de cistos bucais [4]. Este também é um problema complexo, pois a inspeção visual muitas vezes se mostra imprecisa e custosa. Neste sentido, a automatização do processo é muito bem-vinda como uma ferramenta que auxilia o diagnóstico médico, além de permitir que o tratamento possa ser antecipado, aumentando assim as chances de uma boa resposta clínica para o paciente.

Metodologia

Uma rede neural artificial é um método de classificação que se baseia em reconhecer padrões encontrados em bases de dados de treinamento, com o objetivo de generalizar esses padrões bem o suficiente para conseguir categorizar novos dados de forma rápida, automatizada e eficiente. As redes neurais convolucionais foram desenvolvidas para processar dados na forma de matrizes, especialmente imagens, pois possuem um caráter hierárquico, que faz com que as camadas iniciais sejam responsáveis pela extração das características mais básicas da imagem [8].

As últimas camadas de uma rede convolucional profunda também são conhecidas como “classificadores”. Elas recebem os descritores das imagens extraídos nas camadas convolucionais e determinam a classificação mais provável para cada imagem. Neste projeto foram utilizados os seguintes classificadores:

Fully Connected Layer (FCL): fileira simples de neurônios totalmente conectados à camada anterior.

Support Vector Machine (SVM): encontra um hiperplano em um espaço N-dimensional que melhor

separe os pontos que representam os dados [13].

K-Nearest Neighbors (KNN): verifica a classe dos K descritores mais próximos ao descritor da imagem a ser classificada e escolhe a classe mais recorrente entre estes K vizinhos.

Random Forest (RF): constroi múltiplas árvores de decisão e as combina para obter uma classificação mais acurada.

Linear Discriminant Analysis (LDA): procura uma combinação linear de variáveis que melhor separa duas classes [3, 14].

Além desses, outros três classificadores também foram implementados. Eles usam uma estratégia diferente dos anteriores, que consiste em fazer uma combinação de descritores (*feature combination*), em que vetores de características de baixo nível (LBP - *Local Binary Patterns*) foram combinados com vetores de alto nível (redes convolucionais profundas), no intuito de aumentar a acurácia geral da rede [5]. Utilizamos o método LBP para calcular vetores de distâncias entre imagens e classes e os combinamos com os vetores de confiança fornecidos pelos classificadores anteriores. Para fazer essa combinação, foram testados três métodos diferentes, cada um deles resultando em um novo classificador.

SVM + LBP: a combinação foi feita usando as distâncias para calcular pesos para o vetor de confiança.

SVM + LBP + NN: o vetor de confiança foi concatenado com o vetor de distâncias e usado para treinar uma rede neural simples, responsável por determinar a melhor combinação entre os vetores.

SVM + LDA + LBP + NN: mesmo princípio do classificador anterior, mas concatena os vetores de confiança dos métodos SVM e LDA com o vetor de distâncias.

Cada um desses oito classificadores determinou a qual classe cada imagem de teste pertence. A partir disso, foi feito um sistema de votação, conhecido como comitê de máquinas (*ensemble*) [11, 12], em que a escolha de cada classificador representa um voto. O intuito foi aumentar a acurácia final, escolhendo a classe que obteve o maior número de votos.

Resultados e Discussão

Primeiramente, utilizamos a base “KTH-TIPS2-2b” para avaliar o desempenho geral da metodologia implementada neste projeto diante de desafios típicos de classificação de imagens. Em seguida analisamos os resultados da aplicação em imagens biológicas de cistos bucais e superfícies foliares.

Todas as redes foram implementadas em *Python*, usando como base os tutoriais de “*Transfer Learning*” da biblioteca do *Pytorch*. Quanto à arquitetura da rede neural, foram utilizados os modelos pré-treinados, ResNet [7] e AlexNet. As abordagens escolhidas para tratar os parâmetros da rede foram as seguintes:

Ajuste fino (*finetuning*): depois da rede ser pré-treinada na base externa ImageNet, todos os parâmetros são otimizados usando as bases de dados em questão neste projeto.

Extração fixa de características: a rede neural também é pré-treinada na ImageNet, porém as primeiras camadas mantêm os pesos “congelados”, otimizando apenas os pesos das camadas finais nas bases internas.

Um dos problemas em visão computacional que pode ser tratado por redes neurais é a classificação de imagens independentemente das condições sob as quais as fotos foram tiradas. Sabe-se que fatores externos podem alterar drasticamente as características de uma fotografia: a iluminação do ambiente, o ângulo com que a luz incide no material, a distância e a posição do material em relação à câmera, entre inúmeros outros aspectos. No entanto, espera-se que a rede seja capaz de classificar a imagem corretamente, mesmo submetida a essas alterações. Para analisar esse aspecto, utiliza-se a base de dados conhecida como “KTH-TIPS-2b” [10] (*Textures under varying Illumination, Pose and Scale*), que consiste em um conjunto de fotos de texturas provenientes de 11 materiais/classes distintos. Cada classe é dividida em quatro subclasses nas quais as fotos foram submetidas a variações na iluminação, posição ou escala.

Pode-se ver na Tabela 1 que sempre que a taxa máxima de acerto ficou em torno de 70%, esses resultados podem ser considerados satisfatórios, já que eles representam o desempenho da rede diante de imagens submetidas a condições adversas não vistas durante o treino. Isso significa que a metodologia implementada neste projeto foi capaz de fazer uma boa generalização, mantendo uma taxa de acerto alta mesmo diante de imagens que possuem características não presentes na base de treino, como variações de iluminação, posição ou escala.

O presente projeto aplicou também a abordagem computacional desenvolvida combinando redes convolucionais com *ensembles* de classificadores na identificação de espécies de plantas brasileiras, mais especificamente da base 1200Tex [2]. Quando feita por especialistas, a identificação de espécies de plantas costuma focar na análise das flores e frutos, por estes serem órgãos com características de fácil observação e medição. No entanto, essas estruturas nem sempre estão presentes nas plantas, dependendo da espécie, do período do ano ou de outros fatores dificultantes. Diante disso, especialistas passaram a buscar esta informação em outras partes da planta, em especial nas superfícies foliares, devido à sua disponibilidade em qualquer época.

O problema que surge é que as folhas são um dos órgãos vegetais mais heterogêneos, tanto anatomica quanto morfológicamente, havendo variação significativa inclusive quando uma mesma planta é submetida a condições ambientais distintas. Este cenário exige um ferramental computacional avançado e as redes profundas têm demonstrado capacidade para esta tarefa. Estes métodos são capazes de analisar tanto estruturas externas quanto internas da folha e descrever com precisão atributos que não podem ser inferidos pelo olho humano, como padrões geométricos, texturais e de cores por exemplo.

Na Tabela 2 podemos ver os resultados para o problema de classificação de superfícies foliares. Quando utilizamos a estrutura ResNet, a maior taxa de acerto foi obtida pelo comitê de máquinas (*ensemble*), enquanto que no caso da estrutura AlexNet os melhores resultados foram obtidos pelo classificador SVM + LBP + NN. Comparando o desempenho entre as duas estruturas de rede, é possível ver que a ResNet teve um desempenho consideravelmente melhor.

Tabela 1: acurácia média e desvio padrão com texturas

Classificador	ResNet		AlexNet	
	Finetuning	Fixa	Finetuning	Fixa
FCL	73 ± 4	69 ± 4	73 ± 3	73 ± 1
LDA	70 ± 1	68 ± 3	60 ± 3	59 ± 3
SVM	73 ± 3	71 ± 3	61 ± 3	61 ± 2
RF	63 ± 6	60 ± 2	52 ± 5	51 ± 3
KNN	70 ± 6	72 ± 2	62 ± 3	61 ± 2
SVM + LBP	73 ± 2	71 ± 2	68 ± 3	65 ± 2
SVM + LBP + NN	74 ± 3	71 ± 2	70 ± 2	67 ± 2
SVM + LDA + LBP + NN	72 ± 3	71 ± 1	69 ± 3	65 ± 3
ENSEMBLE	75 ± 1	73 ± 2	65 ± 2	62 ± 1

Tabela 2: acurácia média e desvio padrão com folhas

Classificador	ResNet		AlexNet	
	Finetuning	Fixa	Finetuning	Fixa
FCL	87 ± 1	75 ± 1	83 ± 1	82 ± 1
LDA	79 ± 2	47 ± 2	70 ± 2	61 ± 2
SVM	88 ± 1	80 ± 1	73 ± 2	69 ± 2
RF	76 ± 2	56 ± 2	56 ± 4	52 ± 1
KNN	84 ± 1	75 ± 2	67 ± 1	60 ± 1
SVM + LBP	89 ± 1	77 ± 2	73 ± 1	81 ± 1
SVM + LBP + NN	90 ± 1	81 ± 2	84 ± 2	83 ± 2
SVM + LDA + LBP + NN	87 ± 2	79 ± 1	78 ± 2	73 ± 1
ENSEMBLE	91 ± 1	83 ± 1	80 ± 1	73 ± 2

Cistos odontogênicos são uma lesão da mandíbula, recentemente reclassificados pela Organização Mundial da Saúde como carcinogênicos [4]. Podem ser divididos em três grupos: radiculares, queratocistos esporádicos e queratocistos sindrômicos. Seu estudo é de grande importância, sobretudo devido a suas altas taxas de crescimento e recidiva, o que faz com que a identificação do grupo correto ou do tipo de cisto seja fundamental para fins de diagnóstico médico e indicações do tratamento mais adequado a ser aplicado em cada caso.

Embora atributos histológicos que diferenciem tais cistos sejam bem conhecidos por histopatologistas e permitam alguma análise por parte destes pesquisadores, o processo neste caso é manual, sendo assim fortemente sujeito à subjetividade do diagnóstico feito por humanos, além de não permitir que a análise de um grande conjunto de dados seja possível em um intervalo de tempo razoável.

Para esta base, além das variações já utilizadas na base de superfícies foliares, foram feitos alguns outros experimentos relevantes. Considere **k** para a classe “esporádicos”, **s** para “sindrômicos” e **r** para “radiculares”. Lembrando que ainda existe a categoria dos queratocistos, que é composta pela união entre os cistos esporádicos e os sindrômicos. Portanto, além do teste padrão entre as três classes (**k x s x r**), testamos a capacidade da rede de distinguir entre os dois tipos de queratocistos (**k x s**) e entre os radiculares e os queratocistos (**ks x r**).

Os resultados, mostrados na Tabela 3, foram condizentes com o esperado para as três configurações da base de dados, já que a diferença entre radiculares e queratocistos (**ks x r**) é a mais acentuada, sendo, portanto, a classificação mais fácil a ser feita, chegando a uma taxa de quase 100% de acerto quando utilizada a ResNet com ajuste fino.

Tabela 3: acurácia média e desvio padrão com cistos (ResNet à esquerda e AlexNet à direita)

Classificador	k x s		k x s x r		ks x r		k x s		k x s x r		ks x r	
	Finetuning	Fixa	Finetuning	Fixa	Finetuning	Fixa	Finetuning	Fixa	Finetuning	Fixa	Finetuning	Fixa
FCL	82 ± 4	77 ± 4	86 ± 3	76 ± 2	99 ± 1	93 ± 3	71 ± 7	77 ± 4	77 ± 6	79 ± 3	88 ± 9	95 ± 1
LDA	77 ± 7	74 ± 3	80 ± 4	70 ± 3	96 ± 2	90 ± 2	67 ± 6	65 ± 7	51 ± 7	60 ± 5	76 ± 7	80 ± 3
SVM	80 ± 6	73 ± 7	83 ± 4	76 ± 3	97 ± 2	93 ± 3	68 ± 7	65 ± 6	61 ± 6	66 ± 3	81 ± 6	87 ± 3
RF	77 ± 5	69 ± 3	81 ± 4	65 ± 3	96 ± 3	76 ± 3	67 ± 6	68 ± 4	63 ± 8	64 ± 4	81 ± 5	80 ± 3
KNN	77 ± 9	73 ± 5	81 ± 4	74 ± 4	93 ± 4	91 ± 3	59 ± 9	68 ± 4	59 ± 7	65 ± 4	79 ± 5	83 ± 4
SVM + LBP	81 ± 7	74 ± 6	83 ± 4	76 ± 5	96 ± 2	93 ± 3	72 ± 7	67 ± 5	64 ± 7	66 ± 5	81 ± 7	86 ± 3
SVM + LBP + NN	82 ± 5	77 ± 4	84 ± 3	76 ± 4	99 ± 1	94 ± 3	69 ± 9	68 ± 5	62 ± 6	67 ± 3	81 ± 7	88 ± 4
SVM + LDA + LBP + NN	80 ± 5	77 ± 4	83 ± 4	77 ± 4	98 ± 1	94 ± 3	70 ± 6	70 ± 6	59 ± 7	67 ± 4	81 ± 6	85 ± 2
ENSEMBLE	83 ± 6	80 ± 4	86 ± 3	80 ± 3	99 ± 1	95 ± 3	74 ± 8	75 ± 2	69 ± 7	74 ± 3	84 ± 7	90 ± 3

Podemos observar que o desempenho geral dos três classificadores que utilizaram o método de combinação de descritores de baixo e alto nível foi consistentemente superior ao dos demais classificadores. Isso pode ser considerado como um indicio para a corroboração da premissa de que a combinação de vetores de características de diferentes níveis de complexidade aumenta a acurácia na categorização de imagens.

É notável que o método do ajuste fino apresentou resultados mais satisfatórios, o que era esperado, considerando-se que o tipo de imagem analisada é muito específico e, portanto, a base de dados externa ImageNet não teria tanto a acrescentar no desempenho da nossa rede. Por essa razão, quando todos os pesos

são congelados, exceto os da última camada, obtém-se um resultado inferior comparado a quando apenas se inicializa a rede usando-se a base de dados externa e deixa-se que os pesos sejam otimizados a partir dos dados internos.

Conclusões

Neste projeto implementamos diversos classificadores com o intuito de aumentar a acurácia da rede usando um comitê de máquinas. Os resultados obtidos mostraram que essa estratégia de *ensemble* em geral repetiu o desempenho dos melhores classificadores ou mostrou um aumento na taxa de acerto.

Por outro lado, os classificadores que utilizaram o método de combinação de descritores de baixo e alto nível mostraram resultados consistentemente altos em comparação com os demais classificadores. Esses resultados sugerem o desenvolvimento de estudos futuros aprofundando essas técnicas e testando outras possibilidades de combinação de descritores.

A metodologia desenvolvida é altamente promissora, pois não demanda uma grande base de dados para treino e nem um alto poder computacional. Essas características são essenciais para as bases estudadas neste projeto, considerando que essas imagens são de difícil obtenção. Além disso, tanto o uso da combinação de descritores quanto o do comitê de máquinas são muito convenientes, pois possibilitam melhorar os resultados de acurácia sem a necessidade de incluir novas informações externas ou de buscar por outras abordagens completamente diferentes.

Esse sistema computacional desenvolvido é capaz de auxiliar o médico ou outro profissional da saúde na detecção e diagnóstico a partir de biópsias observadas em microscópio, bem como o botânico ou taxonomista na identificação de espécies de plantas brasileiras.

Referências bibliográficas

- [1] Babak Alipanahi, Andrew DeLong, Matthew T. Weirauch and Brendan J. Frey. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *NATURE BIOTECHNOLOGY*, 33(8):831+, AUG 2015.
- [2] Dalcimar Casanova, Jarbas Joaci de Mesquita Sá Junior and Odemir Martinez Bruno. Plant leaf identification using gabor wavelets. *International Journal of Imaging Systems and Technology*, 19(3):236–243, 2009.
- [3] Richard O Duda and Peter E Hart. Pattern recognition and scene analysis, 1973.
- [4] Joao B. Florindo, Odemir M. Bruno and Gabriel Landini. Morphological classification of odontogenic keratocysts using bouligand-minkowski fractal descriptors. *Computers in Biology and Medicine*, 81:1 – 10, 2017.
- [5] JI Forcén, M Pagola, E Barrenechea, and H Bustince. Combination of features through weighted ensembles for image classification. *Applied Soft Computing*, 84:105698, 2019.
- [6] Ian Goodfellow, Yoshua Bengio and Aaron Courville. *Deep learning*. MIT press, 2016.
- [7] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [8] Yann LeCun, Yoshua Bengio and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436, 2015.
- [9] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Fran-cesco Ciampi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken and Clara I. Sanchez. A survey on deep learning in medical image analysis. *MEDICAL IMAGE ANALYSIS*, 42:60–88, DEC 2017.
- [10] P Mallikarjuna, Alireza Tavakoli Targhi, Mario Fritz, Eric Hayman, Barbara Caputo and Jan-Olof Eklundh. The kth-tips2 database. *Computational Vision and Active Perception Laboratory (CVAP)*, Stockholm, Sweden, 2006.
- [11] Ye Ren, Le Zhang and Ponnuthurai N Suganthan. Ensemble classification and regression-recent developments, applications and future directions. *IEEE Computational intelligence magazine*, 11(1):41–53, 2016.
- [12] Lior Rokach. Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1-2):1–39, 2010.
- [13] Bernhard Scholkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2001.
- [14] Jieping Ye, Ravi Janardan and Qi Li. Two-dimensional linear discriminant analysis. In *Advances in neural information processing systems*, pages 1569–1576, 2005.