

Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Computação
Científica (IMECC - Unicamp)

Relatório MS777

Redes Neurais Convolucionais para Reconhecimento de
Texturas

Convolutional Neural Networks for Texture
Recognition

Aluna: Marina Rocha

Orientador: Prof. Dr. João Batista Florindo

Junho de 2019

Resumo

Com a crescente relevância das Redes Neurais Profundas para resolver o problema de classificação de imagens, diversas áreas de aplicação têm surgido, uma delas sendo a categorização de imagens biológicas. O presente relatório aplica estes conceitos na área médica, na discriminação de cistos bucais, e na botânica, para a identificação de espécies de plantas a partir de imagens da superfície foliar. Além disso, é feita uma análise dos resultados obtidos pela rede neural com imagens de texturas sob condições diversas de iluminação, posição e escala. A rede apresentou resultados satisfatórios, mostrando que o cenário é propício para a aplicação deste método à análise de imagens biológicas, já que redes profundas têm se mostrado eficientes em detectar padrões não identificáveis pelo olho humano, de forma muito rápida, robusta e precisa.

Palavras-chave— Rede convolucional profunda; Classificação de texturas; Classificação de imagens biológicas

1 Introdução

Desde o surgimento dos primeiros computadores, as pessoas se perguntavam quando as máquinas se tornariam inteligentes. Em 1957, com a criação das primeiras redes neurais, abriu-se espaço para a busca por algoritmos de aprendizagem de máquina cada vez mais sofisticados. Em 1997, um destes algoritmos ganhou do campeão mundial de xadrez, instaurando-se assim aos poucos como uma poderosa tecnologia, que permeia todos os aspectos da sociedade moderna: desde buscadores *online* até filtros de conteúdos em redes sociais [5] e muitos outros exemplos.

Atualmente, o interesse nas redes neurais vem crescendo significativamente, principalmente com a popularização das 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 [7] e na predição de sequências de DNA-RNA [2].

A metodologia aqui 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 [3]. Este é um problema naturalmente desafiador, uma vez que as estruturas classicamente usadas para este fim são flores e frutos. Ao mesmo tempo, estes órgãos muitas vezes não são facilmente acessíveis na natureza e 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.

Além dessas duas bases de dados, também foi utilizada a base “KTH-TIPS2-a” [8], que consiste em um conjunto de imagens de texturas, provenientes de onze materiais (classes) distintos, como por exemplo algodão e madeira. Para cada uma dessas classes, existem quatro subconjuntos, que correspondem a variações na iluminação, posição ou escala da foto. Essa base foi usada para estudar o resultado da metodologia das redes neurais diante de mudanças específicas nas condições da foto, por meio da implementação de um modelo que alterna os subconjuntos das classes entre os papéis de treinar e testar a rede.

2 Redes neurais profundas

2.1 Redes neurais artificiais

O sistema de redes neurais artificiais é um método que “aprende” a partir de uma base de dados de treinamento, da qual já se sabe o resultado esperado. O modelo utiliza uma função de custo para medir o quanto a rede “erra”, essa função então é minimizada iterativamente, seguindo um procedimento conhecido como “gradiente descendente”. A cada etapa, os parâmetros (pesos) do sistema são atualizados, até atingir algum determinado critério de parada. Por fim, utilizamos esses pesos, os quais são otimizados usando-se o conjunto de treino, em uma parte da base da qual não conhecemos a classificação *a priori* (conjunto de teste), com o intuito de que a rede tenha uma boa generalização e também atribua classificações corretas a este novo conjunto.

A estrutura de uma rede neural basicamente consiste em camadas de “neurônios” (estruturas que realizam processamentos simples), conectados entre si, como em um grafo. Temos a camada de entrada, a(s) intermediária(s) e a de saída. No caso de análise de imagens, as entradas podem ser os valores RGB de cada pixel, as camadas intermediárias são tipicamente responsáveis pela extração de descritores da imagem, através da otimização dos pesos e da aplicação de funções não lineares (ex.: sigmoide, ReLU, etc.), enquanto que as últimas camadas costumam agir como classificadores, fornecendo a classificação final da imagem analisada [5].

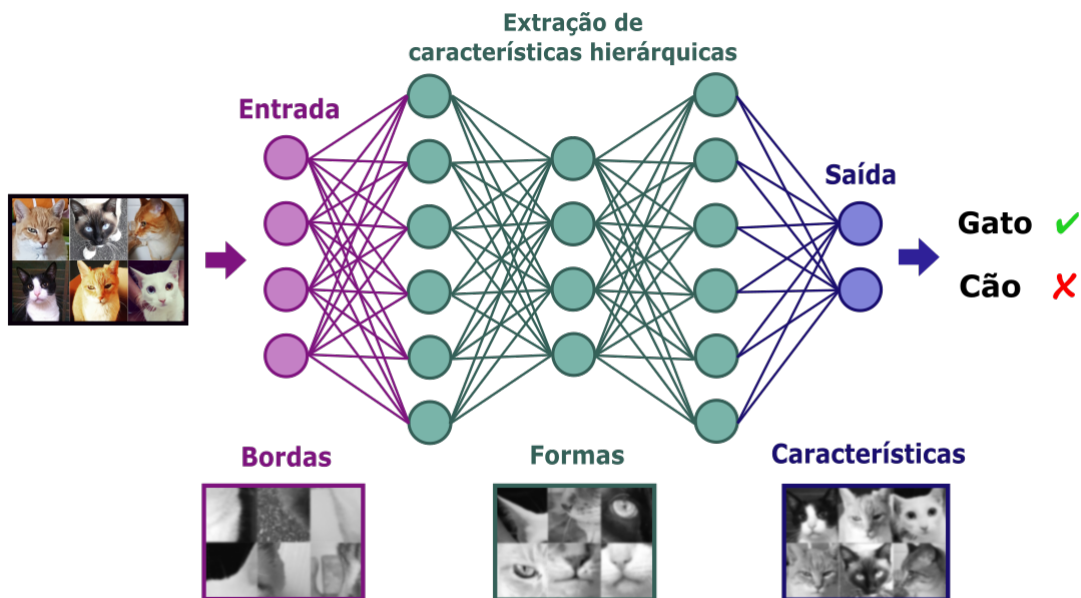


Figura 1: esquema de uma rede neural simples

Expandindo o conceito de redes neurais, temos as redes profundas, que basicamente são redes neurais artificiais que possuem um grande número de camadas. Neste tipo de rede não é possível rastrear o que o algoritmo está fazendo, por isso o método é visto como uma “caixa preta”. No entanto, é este tipo de rede que tem apresentado os resultados mais satisfatórios em análise de imagens e muitas outras aplicações que envolvem dados em larga escala com alta complexidade [9].

2.2 Redes neurais convolucionais

O modelo historicamente mais popular de rede neural é o de perceptrons multicamadas. Porém, por conta de as camadas deste tipo de rede serem totalmente conectadas, o método é propenso a sofrer da conhecida “maldição da dimensionalidade”, ou seja, quanto maior a dimensão do problema, menos viável computacionalmente ele pode se tornar. Além disso, no caso particular de imagens digitais, essas redes não levam em conta a estrutura espacial das imagens, ignorando padrões locais e tratando pixels próximos e distantes da mesma forma (vetorização das imagens).

Nesse contexto surgiram as redes convolucionais, inspiradas no sistema ótico biológico. Elas são formadas por uma estrutura diferenciada, que foi projetada para processar dados na forma de matrizes, especialmente imagens. Este tipo de rede apresenta algumas características principais:

- **Conexões locais (*feature maps*):** nas redes convolucionais, analisamos características locais das imagens, utilizando submatrizes conhecidas como *feature maps*, mantendo a estrutura original da matriz de pixels.
- **Pesos compartilhados:** cada *feature map* possui um conjunto fixado de pesos, que atuam como um filtro sobre a imagem. Esse mesmo filtro é aplicado ao longo de toda a matriz original, sendo apenas deslocado. A ideia por trás dessa propriedade é a de que um filtro otimizado para identificar rostos, por exemplo, deva ser capaz de identificar um rosto em qualquer lugar da imagem.
- **Condensamento (*pooling*):** o papel da camada de condensamento (*pooling layer*) é mesclar características similares, ou seja, a matriz original será dividida em submatrizes e cada uma dessas submatrizes será condensada em um único valor, gerando assim uma nova matriz de dimensão reduzida. Um desses métodos é o *max pooling*, que consiste em escolher o número de maior valor. Outra opção seria tirar a média das entradas da submatriz (*average pooling*).
- **Múltiplas camadas (*deep learning*):** todo esse processo de montar os *feature maps*, otimizar os pesos compartilhados, aplicar uma função não-linear e condensar a matriz é repetido várias vezes ao longo da rede. Isso caracteriza as múltiplas camadas de aprendizagem (*deep learning*). Redes profundas exploram a propriedade de que muitas estruturas naturais possuem uma hierarquia intrínseca, ou seja, características de alto nível de escala (globais) estão também presentes em níveis mais baixos (locais).

Uma das arquiteturas mais comuns de redes convolucionais consiste em algumas camadas convolucionais e camadas de condensamento intercaladas. O papel destas camadas é extrair os descritores dos dados de entrada, enquanto que as últimas camadas servem como classificadores. Estas últimas costumam ser camadas simples de fileiras de neurônios. Por conta do viés hierárquico, nas camadas mais iniciais é possível que se usem até mesmo pesos otimizados em bases de dados externas, que sejam significativamente maiores, já que estas primeiras camadas seriam responsáveis por identificar as características de mais baixo nível, isto é, de caráter mais universal [6].

Uma dificuldade recorrente na aplicação de métodos de aprendizado profundo em geral, e especialmente das redes convolucionais, é a escassez de dados para treinamento. Para amenizar esse problema, usamos técnicas de transferência de conhecimento (*transfer learning*) em todas as implementações apresentadas. Nessa técnica, a rede é pré-treinada em uma base pública de grande porte, no caso, a base ImageNet usando a arquitetura de rede ResNet.

Nesse contexto, utilizaremos duas abordagens distintas. Em uma delas, que chamaremos de “rede com ajuste fino”, em vez de iniciarmos aleatoriamente os pesos da rede, inicializamos utilizando os pesos otimizados pela rede ResNet pré-treinada, e então

deixamos a rede ajustar os pesos normalmente. Na outra, que chamaremos de “rede com extração fixa de características”, congelamos todos os pesos da rede, exceto os da última camada, que farão o papel do classificador das características extraídas pela rede convolucional.

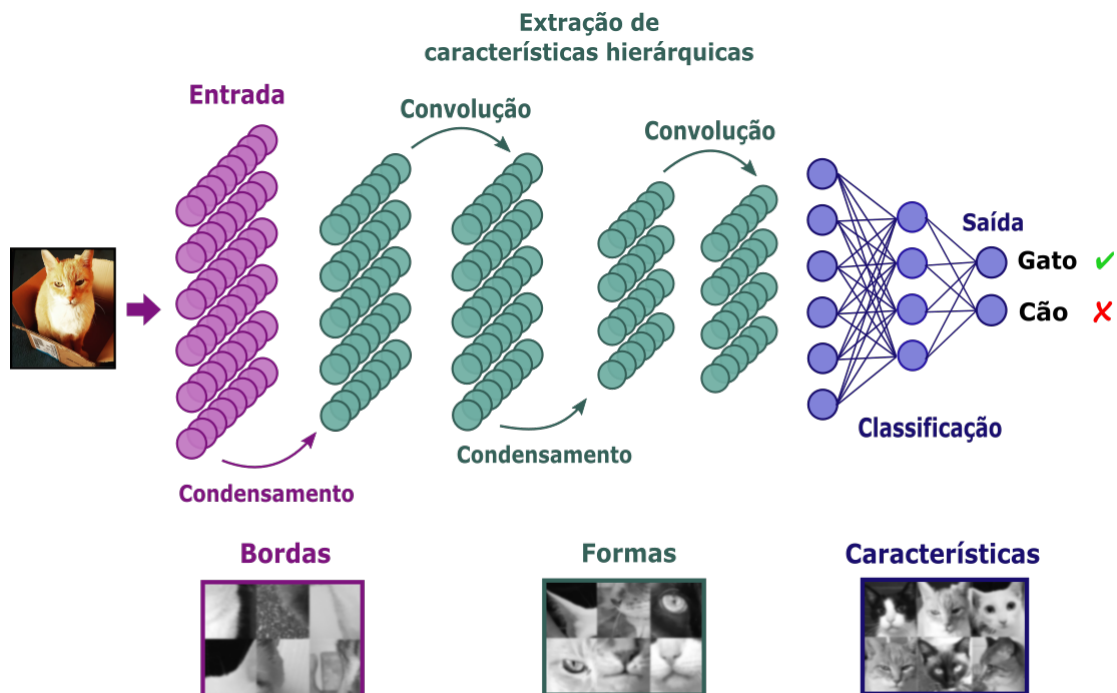


Figura 2: Esquema de uma rede convolucional simples.

3 Aplicações em imagens biológicas

As imagens biológicas (tanto animais quanto vegetais) têm sido uma das áreas mais proeminentes em fornecer dados complexos e em larga escala, apropriados para a análise por redes profundas. Neste cenário, técnicas de aprendizado profundo são propícias para detectar e discriminar padrões escondidos que não são identificáveis por descritores clássicos de imagens e muito menos pelo olho humano.

Um aspecto muito interessante das redes profundas neste cenário é que os atributos mais relevantes da imagem são aprendidos de modo hierárquico: as primeiras camadas detectam características de baixo nível e mais universais enquanto as últimas capturam nuances mais específicas do objeto representado. Curiosamente, este paradigma hierárquico está relacionado com a abordagem multiescala e com diferentes perspectivas, que profissionais como patologistas e taxonomistas já costumam usar há muito tempo.

3.1 Identificação de espécies de plantas a partir de imagens da superfície foliar

O processo de identificação da espécie de um organismo vegetal é de grande importância tanto para especialistas de áreas como botânica, agricultura, fitoterapia, entre outros, quanto para o público leigo. Em termos práticos, uma identificação precisa permite um melhor controle de parâmetros como quantidade e tipo de nutrientes e corretores de solo a serem aplicados, além de outras condições passíveis de serem reguladas, assim aumentando a produtividade de uma safra ou a acurácia de um processo de melhoramento genético, por exemplo. Este estudo também é importante em diversas outras aplicações que envolvam plantas, como as medicinais ou em bioenergia.

Assim como na abordagem convencional, a identificação automatizada de espécies também se baseia na análise de características-chave presentes na planta e que permitem

a distinção entre as diferentes espécies. Classicamente, especialistas costumam focar esta análise em flores e frutos, por estes serem órgãos com características fáceis de serem observadas e medidas e que são capazes de distinguir fielmente entre espécies. Ocorre entretanto que estes órgãos não estão facilmente disponíveis em todas as plantas nem em todas as épocas do ano. Diante disso, especialistas passaram a buscar esta informação em outras partes da planta e em especial nas folhas, devido à sua disponibilidade em qualquer época.

O problema que surge é que as folhas são um dos órgãos vegetais mais heterogêneos tanto anatomicamente quanto morfologicamente, 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.

O presente projeto aplica a abordagem computacional desenvolvida usando redes convolucionais na identificação de espécies de plantas brasileiras, mais especificamente da base 1200Tex proposta e estudada em [3]. Algumas amostras desta base são ilustradas na Figura 3.

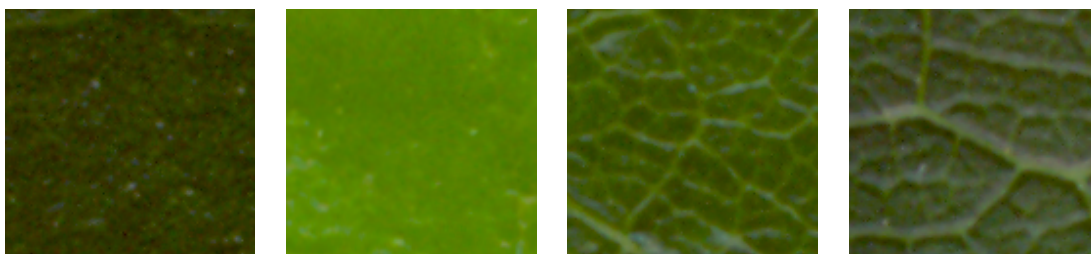


Figura 3: Exemplos de superfícies foliares analisadas.

3.2 Categorização de cistos bucais

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. A Figura 4 ilustra imagens de algumas amostras destes cistos. 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 (se de fato temos um queratocisto ou um cisto radicular, por exemplo) 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 preencher esta lacuna, métodos computacionais de análise de imagens microscópicas vêm sendo aplicados na literatura nos últimos anos, alcançando resultados promissores. Em particular, as redes neurais convolucionais profundas se apresentam como uma abordagem bastante promissora neste tipo de problema [7].

Neste sentido, propõe-se aqui que os descritores extraídos pela metodologia desenvolvida sejam empregados na identificação e discriminação de cistos bucais com base nas imagens microscópicas coletadas a partir da biópsia de tecidos afetados.

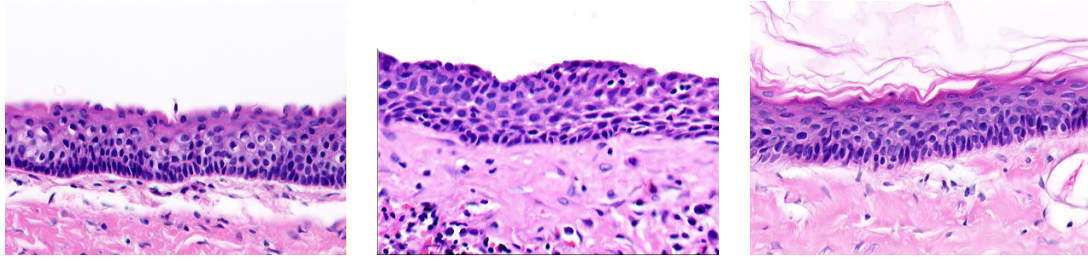


Figura 4: Imagens ilustrativas de algumas amostras dos cistos bucais estudados neste projeto.

3.3 Categorização de texturas sob condições distintas de iluminação, posição e escala

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. Sabemos 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 a 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 da rede neural utilizada neste projeto, utilizamos a base de dados conhecida como “KTH-TIPS2-a” [8] (*Textures under varying Illumination, Pose and Scale*), que consiste em um conjunto de fotos de texturas provenientes de 11 materiais/-classes distintos, sendo eles: folha de alumínio, pão integral, veludo, cortiça, algodão, biscoito, folha de alface, linho, pão branco, madeira e lã. Cada classe, por sua vez, é dividida em quatro subclasses nas quais as fotos foram submetidas a variações na iluminação, posição ou escala.

A metodologia utilizada consiste em escolher uma dessas amostras de cada classe e usar apenas esse conjunto de imagens para testar a rede, enquanto as outras três subclasses são usadas para treinar a rede. Com isso, podemos analisar qual é o resultado da rede ao se deparar com um conjunto de teste cujas imagens estão submetidas a alguma condição de iluminação, posição ou escala à qual ela não tinha sido exposta anteriormente. O processo é repetido até que todas as subclasses tenham sido utilizadas uma vez como conjunto de teste da rede.

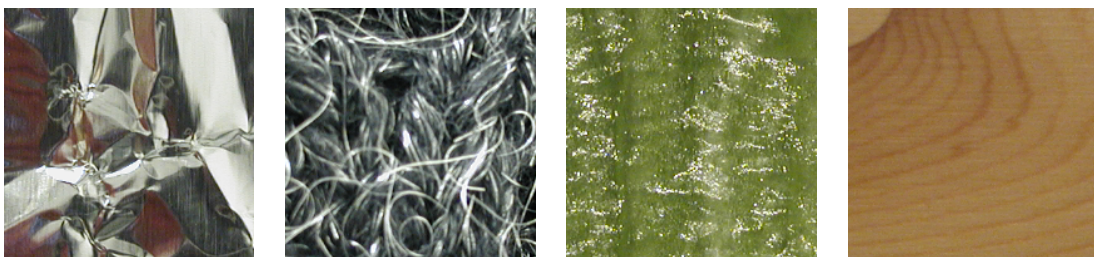


Figura 5: Exemplos de texturas analisadas.

4 Resultados e Análises

Todas as redes foram implementadas em *Python*, usando como base os tutoriais de “*Transfer Learning*” [1] da biblioteca do *Pytorch*¹.

¹Os códigos podem ser encontrados em <https://github.com/MarinaRocha29/MS777>

4.1 Identificação de espécies de plantas

Os conjuntos de teste e treino foram separados aleatoriamente, numa proporção de 20 fotos para o teste e 40 para o treino, para cada uma das 20 classes. Abaixo constam os resultados aproximados da acurácia máxima e a matriz de confusão obtidos usando as duas abordagens de pré-treinamento citadas anteriormente:

- Rede com ajuste fino: 92,50%

11	0	0	1	4	1	0	2	1	0	0	0	0	0	0	0	0	0	0	0
0	19	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	17	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1
1	0	0	0	16	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2
0	0	0	0	0	17	0	2	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	2	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	2	0	0	0	0	15	0	0	0	2	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0
0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	18	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20

- Rede com extração fixa de características: 71,25%

10	0	0	0	5	0	0	1	1	1	0	0	0	0	0	1	0	1	0	0
0	18	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	11	0	0	0	1	0	0	0	1	0	2	1	0	2	0	0	1
2	0	0	0	11	0	0	0	3	1	0	1	0	0	1	0	0	0	1	0
0	0	0	0	1	12	0	2	0	0	0	2	0	0	0	2	1	0	0	0
0	1	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	1	2	0	9	0	0	0	0	0	0	2	0	4	0	0	0
0	0	0	0	5	0	0	0	13	0	0	1	0	0	0	0	0	0	1	0
1	0	0	0	1	0	1	0	0	6	0	1	0	3	6	0	1	0	0	0
0	0	0	0	0	0	2	0	0	0	18	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	14	5	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	5	15	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	2	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	19	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0
1	0	0	2	0	8	2	0	0	0	0	1	0	0	0	1	0	5	0	0
1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	18	0
0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	9

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 congelamos todos os pesos, exceto os da última camada, obtivemos um resultado inferior comparado à quando apenas inicializamos a rede usando a base de dados externa e deixamos que os pesos sejam otimizados a partir dos dados internos.

As matrizes de confusão evidenciam ainda mais a diferença entre a performance das duas metodologias. Na primeira, podemos ver uma clara predominância nos valores da diagonal principal, ou seja, no número de imagens classificadas corretamente, sendo que não há valores significativos fora da diagonal. Na segunda, ainda é possível notar uma predominância na diagonal, mas consideravelmente menos acentuada, além de ser possível encontrar valores significativos fora dela.

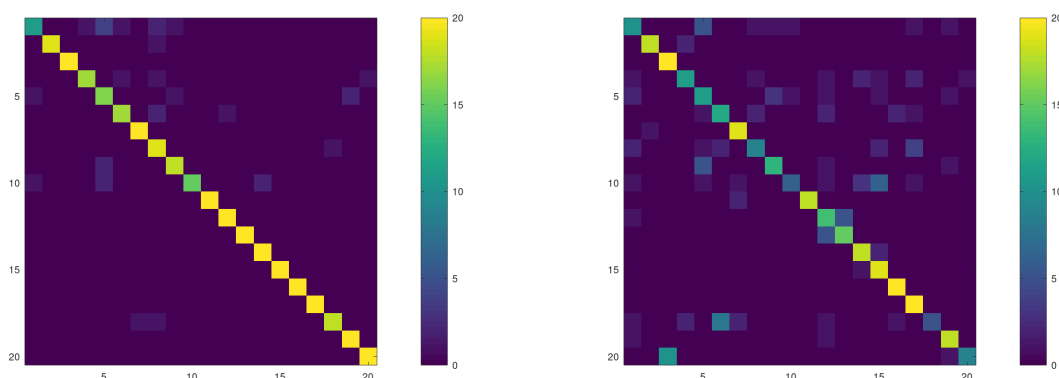


Figura 6: Representação em escala de cores das matrizes de confusão.

Para ilustrar, podemos observar a última linha da segunda matriz de confusão, ela representa a distribuição das classificações atribuídas às imagens pertencentes à classe vinte. Podemos observar que apenas nove imagens foram classificadas corretamente, enquanto que outras dez foram atribuídas erroneamente à classe três.

4.2 Categorização de cistos bucais

Os conjuntos de teste e treino foram separados aleatoriamente, numa proporção de aproximadamente um terço das fotos para o teste e o restante para o treino, para cada uma das três classes. Abaixo constam os resultados aproximados da acurácia máxima e a matriz de confusão obtidos usando as duas abordagens de pré-treinamento citadas anteriormente:

- Rede com ajuste fino: 82,61%

$$\begin{bmatrix} 15 & 0 & 6 \\ 1 & 13 & 1 \\ 0 & 0 & 10 \end{bmatrix}$$

- Rede com extração fixa de características: 78,26%

$$\begin{bmatrix} 18 & 0 & 3 \\ 2 & 12 & 1 \\ 4 & 0 & 6 \end{bmatrix}$$

Podemos inferir desses resultados as mesmas análises feitas para a identificação de espécies de plantas na seção anterior, levando em consideração que por termos uma quantidade consideravelmente menor de imagens e classes, a matriz de confusão não acrescenta tanta informação às análises quanto no caso anterior.

4.3 Categorização de texturas

Para cada classe da base “KTH-TIPS2-a”, temos quatro subclasses, denominadas “*sample a*”, “*sample b*”, “*sample c*” e “*sample d*”. Abaixo constam os resultados aproximados da acurácia máxima obtida usando cada uma dessas amostras e um arranjo aleatório como conjunto de teste da rede.

- Usando “*sample a*” como conjunto de teste
 - Rede com ajuste fino: 79,29%
 - Rede com extração fixa de características: 75,51%
- Usando “*sample b*” como conjunto de teste
 - Rede com ajuste fino: 79,88%
 - Rede com extração fixa de características: 76,68%
- Usando “*sample c*” como conjunto de teste
 - Rede com ajuste fino: 86,36%
 - Rede com extração fixa de características: 87,63%
- Usando “*sample d*” como conjunto de teste
 - Rede com ajuste fino: 78,74%
 - Rede com extração fixa de características: 74,23%
- Usando um arranjo aleatório de imagens como conjunto de teste
 - Rede com ajuste fino: 98,89%

$$\begin{bmatrix} 101 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 107 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 108 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 109 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 96 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 101 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 109 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 100 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 107 & 1 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 107 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 109 \end{bmatrix}$$

- Rede com extração fixa de características: 90,23%

$$\begin{bmatrix} 97 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 3 \\ 0 & 99 & 0 & 3 & 0 & 2 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 95 & 0 & 5 & 0 & 0 & 2 & 2 & 1 & 4 \\ 0 & 0 & 0 & 109 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 84 & 0 & 0 & 7 & 0 & 1 & 3 \\ 0 & 2 & 0 & 8 & 0 & 90 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 3 & 0 & 98 & 0 & 3 & 0 & 3 \\ 0 & 0 & 1 & 0 & 6 & 0 & 0 & 90 & 0 & 0 & 4 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 107 & 0 & 0 \\ 0 & 0 & 5 & 3 & 1 & 0 & 0 & 0 & 0 & 100 & 0 \\ 0 & 0 & 4 & 0 & 20 & 0 & 0 & 1 & 0 & 0 & 84 \end{bmatrix}$$

Podemos ver que sempre que uma amostra foi usada como conjunto de teste, a taxa de acerto ficou em torno de 80%, enquanto que quando usamos um arranjo aleatório de imagens de todas as amostras, o desempenho se mostrou consideravelmente superior. Isso já era previsto, pois no segundo caso, a rede já tinha tido contato com imagens similares (pertencentes à mesma amostra) durante o treino.

Esses resultados usando as amostras como conjunto de teste 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 o método é 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, por exemplo.

As mesmas análises feitas nas seções anteriores podem se aplicar a este caso. Podemos, por exemplo, analisar o valor da quinta coluna da última linha, na segunda matriz de confusão, em que houve um número significativo de imagens que pertenciam à classe “lã”, mas foram atribuídos à classe “algodão”. Esse resultado é compreensível quando analisamos visualmente as imagens dessas duas classes, já que algumas amostras são de fato muito parecidas, sendo difícil a classificação até mesmo para um ser humano.

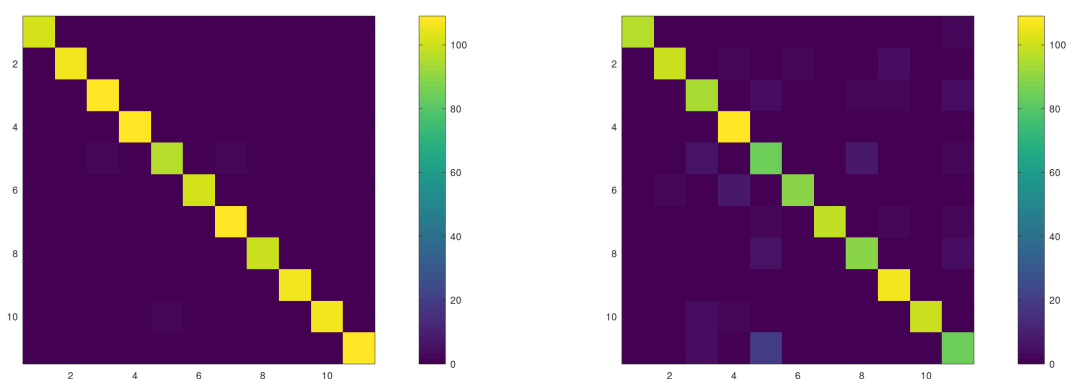


Figura 7: Representação em escala de cores das matrizes de confusão.

5 Conclusão

O projeto abordou o desenvolvimento científico e tecnológico em visão computacional e aprendizado de máquinas, bem como a aplicação na área médica e na botânica. Analisamos a eficiência das redes neurais em categorização de imagens sob condições variadas, mostrando que a rede apresenta um desempenho satisfatório mesmo diante de condições adversas não observadas no processo de treinamento.

Assim, foram investigados aspectos teóricos da área e ao final foi desenvolvido um sistema computacional 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

- [1] Pytorch tutorial. https://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html. Accessed: 2019-06-04.
- [2] 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.
- [3] 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.
- [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] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [6] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.
- [7] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, 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.
- [8] 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.
- [9] Juergen Schmidhuber, Xi Zhen-yuan, Wang Chang-li, Chen He, and Wu Jing. Deep learning in neural networks: An overview. In *2015 International Conference on Computers, Communications, and Systems (ICCCS)*, pages 85–117, 2015.