

Técnicas de Multi-Análise em Redes Neurais Convolucionais Para Classificação de Imagens de Granito

Anselmo Ferreira *

Gilson Girardi †

24 de Outubro de 2016

Resumo

Padronizar técnicas de classificação de imagens é uma tarefa difícil por duas razões (1) a natureza heterogênea de diferentes tipos de imagem, o que exige uma técnica diferente exclusiva para cada tipo e (2) a grande quantidade de classes para um mesmo tipo de imagem, o que pode confundir um classificador tradicional. Nesse primeiro relatório iremos reportar os primeiros resultados obtidos em nosso projeto de pesquisa focado em amenizar esses problemas. Propomos a criação de descritores baseados em Multi-Análise e diferentes redes Neurais Convolucionais aplicadas na mesma imagem. Mostramos nesse primeiro relatório resultados preliminares promissores em classificação de imagens de granito.

Palavras-chave: Multi-Análise; Redes Neurais Convolucionais; Classificação de Imagens.

1 Introdução

O uso abundante de novas tecnologias para a geração de imagens nos dias atuais exige o desenvolvimento de metodologias para a construção de algoritmos de reconhecimento de imagens para diversas aplicações como, por exemplo, auxílio de diagnóstico de doenças por processamento de imagens médicas, reconhecimento de objetos, investigação de documentos de ética duvidosa, entre outros.

Nesse ínterim, um tema central para este trabalho é a utilização de técnicas que aliam conceitos de Visão Computacional e Aprendizado de Máquina. Essas técnicas inicialmente descrevem as imagens via modelos matemáticos, com o objetivo de caracterizar os diferentes padrões presentes na imagem digital. Finalmente, classificadores recebem esses dados numéricos, discriminando representações matemáticas de diferentes objetos através de um treinamento, podendo ser este supervisionado ou não.

*Laboratório Nacional de Computação Científica Petrópolis, RJ, Brasil. anselmo.ferreira@gmail.com

†Laboratório Nacional de Computação Científica Petrópolis, RJ, Brasil. gilson@lncc.com.br

No entanto, a maioria das técnicas de Visão Computacional e Aprendizado de Máquina propostas na literatura científica para classificação de imagens digitais é baseada em fortes suposições prévias sobre o comportamento dos padrões presentes nas imagens, fundamentadas na ideia de que estes padrões não evoluem dinamicamente à medida que novos dados aparecem e na premissa de cenários restritos, com um número pré-definido de classes de treinamento, não se admitindo a possibilidade de que um novo exemplo seja de uma classe desconhecida da base de treinamento. Dessa forma, a aplicação dessas metodologias a um cenário de mundo real fica severamente comprometida.

Neste primeiro relatório, iremos reportar os resultados obtidos em nosso projeto de pesquisa que busca atuar nessas duas questões, propondo para tal novas abordagens de reconhecimento de imagens por Redes Neurais Convolucionais (RNCs) via Multi-Análise. Projetamos um descritor de imagens baseado em RNCs por Multi-Análise (FERREIRA et al., 2016; FERREIRA; ROCHA, 2014; FERREIRA; SANTOS; ROCHA, 2016; FERREIRA et al., 2015a; FERREIRA et al., Submetido; FERREIRA, 2016). Multi-Análise é uma técnica proposta pelo candidato em seu doutorado (FERREIRA, 2016), em que diversos cenários de investigação das imagens são considerados para a sua classificação. Esses diversos cenários envolvem a análise em múltiplas resoluções na imagem, na multi-direcionalidade da vizinhança dos pixels, na multi-perturbação das imagens para ressaltar artefatos, na multi-representação da mesma imagem, entre outros.

A primeira aplicação de RNCs e Multi-Análise nessa pesquisa envolveu o uso do cenário de *Múltiplos Dados*, onde a imagem de entrada é dividida em blocos que serão individualmente classificados. Também propomos um novo cenário, denominado *Multi-Configuração*, onde diferentes Redes Neurais Convolucionais com diferentes parâmetros e arquiteturas são utilizadas e as suas descrições ou classificações são fundidas. Pretendemos atuar em diversas aplicações para demonstrar a variedade do uso das técnicas desenvolvidas nesse projeto. Neste primeiro relatório iremos reportar os resultados iniciais obtidos em classificação de imagens de rochas de granito.

2 Classificação de Imagens de Granito

Na indústria de construção o granito tornou-se cada vez mais popular devido a sua combinação de força, beleza e preço relativamente acessível. Existem diversos tipos comerciais de granito, que diferem tanto em cor quanto em textura. No entanto, as suas denominações podem frequentemente mudar de um país para outro. Por causa de sua origem natural, a aparência visual de granitos com o mesmo conteúdo mineralógico pode diferir significativamente. Com a massiva popularização do uso do granito (o mesmo responde por 60% da produção mundial de produtos crus e acabados (Internazionale Marmi e Macchine Carrara S.p.A., 2011)), o uso de inspeção visual por peritos torna-se cada vez mais custosa e pode levar a atrasos, o que exige a construção de técnicas automatizadas de reconhecimento de tipo de granitos.

Diversos trabalhos anteriores focaram na classificação dos padrões de diferentes tipos de granitos, onde o uso de análise de textura é o mais utilizado para a classificação. Kurmyshev et al. (KURMYSHEV; SÁNCHEZ-YÁÑEZ; FERNÁNDEZ, 2013) descreveram uma aplicação baseada em Representação de Clusters Coordenados para o controle de qualidade de ladrilhos de granito polido do tipo ‘Rosa Porriño’. Bianconi e Fernández (BIANCONI; FERNÁNDEZ, 2006) empregaram diferentes filtros de Gabor para classificar diferentes tipos de granito em um canal de cor apenas e Lepistö et al. (LEPISTÖ;

KUNTTU; VISA, 2005) usaram uma abordagem similar, no entanto em cada canal de cor separadamente.

Em outro grupo de métodos os dados são adquiridos por espectrofotometria, que diferentemente da aquisição tradicional de imagens, captura também os comprimentos de onda ultravioleta próximo e infravermelho próximo. A abordagem desenvolvida por Araújo *et al.* (Araújo María *et al.*, 2010) pertence a esse grupo. Nesse trabalho os autores empregam um espectrofotômetro de contato para capturar dados do espectro em áreas aleatoriamente escolhidas e os dados coletados são usados em um classificador SVM. Finalmente existem as técnicas que usam informações de cores e textura simultaneamente para a classificação de granitos, sendo um exemplo desse ramo de técnicas o estudo conduzido por Bianconi *et al.* (BIANCONI *et al.*, 2012).

3 Conceitos Básicos

Nesta Seção apresentamos os conceitos básicos para melhor entendimento da técnica proposta, que se vale de fusões via multi-análise de Redes Neurais Convolucionais. Apresentamos primeiramente a multi-análise e depois discutimos conceitos de Redes Neurais Convolucionais nas próximas subseções.

3.1 Classificação de Imagens por Multi-Análise

A multi-análise é um conjunto de técnicas de descrição e classificação de imagens proposta pelo candidato em seu doutorado (FERREIRA, 2016). Inicialmente aplicada em Análise Forense de Documentos, essa abordagem é composta pela fusão de técnicas convencionais de descritores ou por fusões de descritores convencionais, usando para tal uma ou mais disciplinas (cenários investigativos) aplicados na imagem investigada. Tais disciplinas são descritas a seguir:

1. Multi-Direcionalidade: a vizinhança do pixel é analisada na tarefa de classificação.
2. Multi-Escala: diversas dimensões da imagem de entrada são consideradas na investigação de padrões ou artefatos.
3. Multi-Representação: diversas transformações são realizadas na imagem de forma a ressaltar padrões ou artefatos de interesse para a aplicação.
4. Multi-Dados: diferentes áreas de interesse da imagem são analisadas.
5. Multi-Perturbação: é executado um procedimento na imagem de forma a ressaltar o artefato (caso exista) na imagem investigada.

Tais cenários originaram uma série de técnicas de classificação utilizadas no doutorado do candidato, mais especialmente na área de computação forense. Por exemplo, uma fusão convencional por voto de descritores baseados em multi-perturbação e multi-escalas formou uma solução para a detecção de imagens manipuladas por filtragem mediana (FERREIRA; ROCHA, 2014; FERREIRA; SANTOS; ROCHA, 2016), outra fusão convencional por concatenação de vetores de característica de descritores baseados em multi-direcionalidade e multi-escala foram aplicados em detecção de impressora laser fonte de um documento (FERREIRA *et al.*, 2015a) e outra solução, de mesma aplicação, usou voto de descritores baseados em RNCs usando cenários de multi-representações e multi-dados

(FERREIRA et al., Submetido). Finalmente, uma nova técnica de fusão de classificadores foi proposta, usando conceitos de multi-direcionalidade e multi-escalas de descritores convencionais para detectar manipulações de imagem por cópia e colagem (FERREIRA et al., 2016).

O uso da multi-análise apresenta diversas extensões, conforme discutido em (FERREIRA, 2016). A primeira é a possibilidade de criar novas combinações dos cenários existentes para a criação de novos descritores para tarefas de classificação de imagem, já que os cenários podem ser também utilizados para ressaltar padrões ao invés de simplesmente artefatos. A segunda é a flexibilidade da técnica de aceitar novos cenários com a possibilidade de fundi-los ou não com os cenários existentes de forma a criar novos descritores de imagem. Iremos mostrar nesse relatório os primeiros avanços em criar um novo cenário de multi-análise baseado em fusões de RNCs aplicados a classificação de imagens fora do âmbito de computação forense, ou seja, em imagens de rochas do tipo granito. Assim, demonstramos o potencial da técnica para geração de descritores de imagens de propósito geral.

3.2 Redes Neurais Convolucionais

Redes Neurais Convolucionais (RNCs) têm se mostrado atrativas desde o ano de 2012 (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), em tarefas complexas de classificação de imagens. Inicialmente propostas por LeCun *et al.* (LECUN et al., 1998), trata-se da aplicação de aprendizado profundo (*deep learning*), em que Redes Neurais de muitas camadas são aplicadas em imagens. Diferentemente de Redes Neurais comuns, a principal operação que ocorre nas imagens é a chamada *convolução*. Uma imagem gerada pela convolução de um filtro com outra imagem é definida por:

$$o(m, n) = \sum_{i=1}^q \sum_{j=1}^q f(i, j) I(m - i, n - j) \quad (1)$$

onde:

o é a imagem de saída da convolução

q é a dimensão da máscara de convolução, assumindo que a máscara seja quadrada. Caso $q=3$ a máscara será 3×3

i, j são as posições em duas dimensões da máscara de convolução

m, n são as posições em duas dimensões da imagem de entrada e de saída da convolução

$f(i, j)$ é o peso presente na posição (i, j) da máscara de convolução

$I(m, n)$ é o valor do pixel na imagem

As camadas básicas de RNCs são listadas abaixo:

1. **Camada de Entrada** (ou *Input Layer*): Onde a imagem é inserida na rede. Os dados de entrada podem ser os pixels ou pixels transformados, o que pode enfatizar melhor alguns aspectos específicos das imagens.

2. **Camada de Convolução** (ou *Convolutional Layer*): contém diversos filtros lineares de tamanho fixo usados para realizarem a convolução, gerando o chamado *mapa característico* da imagem. Tais filtros podem ressaltar padrões úteis na caracterização da imagem como, por exemplo, bordas, padrões regulares, entre outros.
3. **Camada RELU** (ou *RELU Layers*): Normalmente ocorrem após a operação de convolução e aplicam uma função não linear à saída x da camada anterior, como por exemplo, $f(x) = \max(0, x)$. De acordo com Krizhevsky et al. (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), elas são usadas para a rede convergir mais rápido.
4. **Camada de Subamostragem** (ou *Subsampling Layers* ou *Pooling Layers*): Essas camadas garantem que a rede foca apenas nos padrões mais importantes gerados pelas camadas convolucionais. Essa camada sumariza os dados ao deslizar uma janela nos mapas de características gerados pelas camadas convolucionais, aplicando operações nos dados da imagem presentes na janela como, por exemplo, média ou máximo dos valores dentro da janela que desliza pela imagem, criando assim um novo mapa de características. Desta forma reduz-se a dimensionalidade dos mapas de característica a serem usados pelas seguintes camadas.
5. **Camadas totalmente conectadas** (ou *Fully Connected Layers*): Usadas para o reconhecimento dos padrões gerados pelas camadas anteriores. Se localizam ao fim da rede agindo como classificadores, geralmente são do tipo *soft-max* usados para determinar a classe associada à imagem de entrada.

A Figura 1 mostra uma possível arquitetura de uma RNC. Tipos e arranjos de camadas variam dependendo da aplicação-alvo.

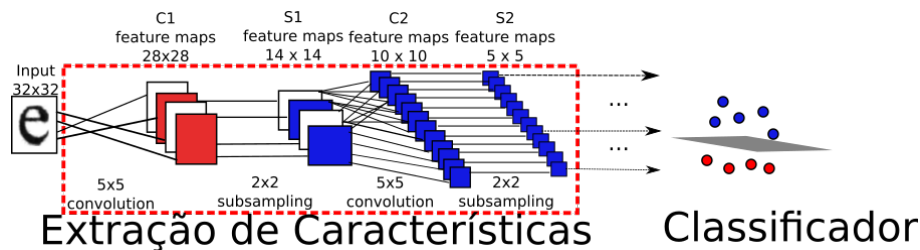


Figura 1 – Arquitetura comumente utilizada em Redes Neurais Convolucionais. A imagem de entrada é transformada em mapas de características pela primeira camada convolucional $C1$. Um estágio de subamostragem $S1$ reduz as dimensões desses mapas e os processos são repetidos nas camadas $C2$ e $S2$. Finalmente o classificador é treinado com os dados gerados pela última camada, que normalmente é uma camada convolucional precedida por uma camada RELU.

O principal atrativo das RNCs é a quantidade reduzida de parâmetros (pesos dos filtros) a serem aprendidos. As camadas convolucionais das RNCs feitas de máscaras de tamanhos reduzidos permitem uma extração eficiente de características de alto nível que são utilizadas pelas camadas totalmente conectadas. O treino de uma RNC é por propagação reversa do erro via descida do gradiente estocástico. Erros de classificação na etapa de treinamento são usados para atualizar os pesos das máscaras de convolução das camadas convolucionais e das camadas totalmente conectadas. Em algumas aplicações usam-se redes já treinadas apenas para extrair características de imagens, removendo para tal a camada totalmente conectada e usando os vetores de característica em outro

classificador, processo esse denominado transferência de aprendizado (ou *transfer learning*) (JOHNSON; KARPATY, 2016; YOSINSKI et al., 2014).

Apesar de poderosas em representar padrões presentes nos dados, dois principais problemas ocorrem nas RNCs: (i) redes muito profundas (com muitas camadas) são necessárias para uma boa classificação e (ii) muitos dados (imagens) são necessários para o treino dessas redes. Em nossa técnica proposta tratamos ambos os problemas ao primeiramente criar redes não tão profundas (com menos de 20 camadas), mas que podem funcionar como uma profunda no âmbito da boa performance esperada de classificação. Finalmente, resolvemos o problema da falta de muitas imagens para a aplicação descrita utilizando o cenário de *Multi-Dados*. Ou seja, dividimos cada imagem em pequenos blocos (*patches*), gerando desta forma milhares de imagens para o treino da rede. Com isto, podemos usar a técnica de votação majoritária dos blocos para classificar uma imagem. Discutiremos a técnica proposta na próxima Seção.

4 Técnica proposta

Nesta pesquisa propomos o uso do cenário denominado *Multi-Configuração*, o qual possui dois elementos básicos: o cenário de *Multi-Dados* e as RNCs. Para tal, aplicamos os mesmos dados de entrada, representados por diferentes regiões de interesse da imagem (cenário de *Multi-Dados*) de tamanho 28×28 em três diferentes Redes Neurais Convolucionais com diferentes arquiteturas, cenário este portanto criado e denominado *Multi-Configuração* nesta pesquisa. Assim, cada imagem de entrada, convertida em tons-de-cinza, é dividida em blocos, gerando o cenário de *Multi-Dados*. Em seguida, cada bloco será utilizado como entrada de diferentes RNCs, cenário este denominado *Multi-Configuração*. A Figura 2 detalha o funcionamento da técnica proposta.

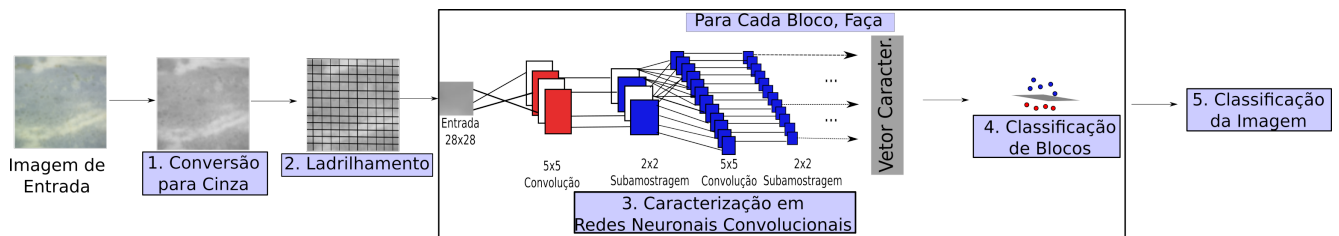


Figura 2 – Desenho ilustrando os passos da técnica proposta para caracterização e classificação de imagens de granito.

Descreveremos nas próximas subseções cada passo da técnica isoladamente.

4.1 Conversão para Cinza

Nessa etapa a imagem de entrada é convertida para nível de cinza, excluindo a informação de cor na análise. Dada uma Imagem de entrada com 3 canais RGB, sua conversão para imagem em tons de cinza C ocorre através da fórmula:

$$C = 0.2989 * R + 0.5870 * G + 0.1140 * B \quad (2)$$

4.2 Ladrilhamento

Após a conversão da imagem para nível de cinza a mesma é subdividida em pedaços 28×28 , sendo que apenas os pedaços originais da imagem (fora das bordas ou sem preenchimento artificial de valores) são usados. Tal procedimento é útil pelas seguintes razões (i) gerar mais dados para o treino das Redes Neurais sem o uso do aumento artificial de dados e (ii) se valer da vantagem da votação majoritária da classificação dos blocos de uma imagem de granito de teste e (iii) podemos utilizar RNCs em imagens de alta resolução, já que, por questões de complexidade computacional, as RNCs somente aceitam imagens com baixa resolução. Estamos nesta etapa aplicando o cenário de *Multi-Dados* (diversas regiões de interesse) na investigação.

4.3 Caracterização em Redes Neurais Convolucionais

Na fase de treinamento, utilizamos os blocos gerados na etapa anterior como entrada de Redes Neurais Convolucionais, onde a rede será treinada e seus pesos de filtros serão estimados. Em seguida, os blocos das imagens de treino e de teste são utilizados como entrada das RNCs obtidas no treinamento, porém, extraíndo a camada totalmente conectada, o que faz a saída de cada rede ser apenas um vetor de características usados para treino e teste de outro classificador. Usamos três redes que possuem arquiteturas distintas, mas aplicadas aos mesmos dados de entrada. As arquiteturas dessas redes são descritas a seguir:

- Rede MNIST1: baseada na arquitetura utilizada no desafio de reconhecimento de dígitos da base MNIST ([VLFEAT, 2016](#)), ela atua em imagens de entrada 28×28 e é composta de oito camadas: quatro camadas Convolucionais, duas de Subamostragem, uma RELU e uma Totalmente Conectada.
- Rede MNIST2: estendemos a rede MNIST1, criando uma nova rede composta de dez camadas: cinco camadas Convolucionais, três de Subamostragem, uma RELU e uma Totalmente Conectada.
- Rede MNIST3: estendemos ainda mais a rede MNIST1, criando uma nova rede composta de doze camadas: seis camadas Convolucionais, quatro de Subamostragem, uma RELU e uma Totalmente Conectada.

Cada uma das redes, irá gerar, respectivamente, vetores de características de 500, 256 e 256 dimensões respectivamente. Dentro do cenário de *Multi-Configuração* proposto, usamos o resultado das três redes simultaneamente para o teste, usando duas metodologias:

1. Fusão por voto (FUSAO_VOTO): a classificação final de um bloco será por votação majoritária da classificação por florestas aleatórias dos vetores de características gerados pelas três redes para esse bloco.
2. Fusão por concatenação (FUSAO_CONCAT): classifica-se por florestas aleatórias o vetor de características construído ao fundir vetores de características das três redes. Concatenando tais vetores temos um vetor de características final de $500+256+256=1012$ dimensões.

Ao usarmos a fusão por concatenação (FUSAO_CONCAT), o próximo passo do algoritmo é normalizar (escalonar) os vetores de características concatenados, fazendo

com que os vetores gerados possuam alguma propriedade de interesse. Existem diversas maneiras de se normalizar um vetor de característica. Escolhemos a forma mais simples, que é dividir o vetor pela sua norma. Dado um vetor de características \vec{V} , sua norma p é calculada como:

$$\|\vec{V}\|_p = \left(\sum_{i=1}^n |\vec{V}(i)|^p \right)^{\frac{1}{p}} \quad (3)$$

onde i é o i -ésimo elemento do vetor e n é o número de elementos do vetor.

Em nosso caso utilizamos $p = 2$, norma esta denominada *Norma Euclidiana*. O vetor de característica final $\vec{V}f$ é, portanto

$$\vec{V}f = \frac{\vec{V}}{\|\vec{V}\|_2} \quad (4)$$

desta forma nós escalonamos os componentes do vetor de características de forma que o comprimento (ou magnitude) de cada vetor seja sempre um.

4.4 Classificação de Blocos

Tanto nas etapas de treino e de teste, aplicamos os vetores gerados pelas redes pré-treinadas em um classificador de padrões. Escolhemos o classificador por *florestas aleatórias* por dois motivos (i) trata-se na verdade de diversos classificadores por árvores de decisão, onde a classificação final é dada por voto de cada classificador individual e (ii) em trabalhos anteriores (CARUANA; NICULESCU-MIZIL, 2006; LIU et al., 2013; FERNÁNDEZ-DELGADO et al., 2014) este classificador mostrou melhor performance para tarefas de classificação multiclasse de imagens de diversos propósitos se comparado a outros classificadores.

4.5 Classificação da Imagem

Após a classificação individual de cada bloco, iremos classificar a imagem analisando seus blocos para identificar a classe mais predita pelo classificador. Dado um vetor \vec{x} que contém a classificação de cada um dos b blocos de uma imagem, a classe final de uma imagem de granito G é:

$$classe(G) = moda(\vec{x}) \quad (5)$$

onde $moda(\vec{x})$ é o valor (classe predita) mais frequente no vetor \vec{x} .

5 Experimentos

Nessa Seção apresentaremos os resultados encontrados em experimentos iniciais, realizados em um conjunto de validação. A seguir, apresentaremos os materiais e métodos utilizados e finalmente discutiremos sobre os resultados encontrados.

5.1 Materiais e Métodos

Antes de mostrarmos os resultados propriamente ditos, iremos nessa Subseção apresentar os materiais e métodos utilizados para validarmos nossa técnica proposta em

comparação a algumas técnicas estado da arte. Discutiremos a seguir a base de imagens de granito utilizada, a metodologia e métricas usadas para a comparação, os parâmetros usados na técnica proposta e o estado da arte utilizado.

5.1.1 Base de Imagens

A base de dados utilizada nos experimentos é a mesma usada no trabalho de *Bianconi et al.* (BIANCONI et al., 2015). Seu conteúdo é de 1000 imagens RGB de dimensões 1500×1500 , sendo essas imagens subdivididas em 25 classes com 40 imagens por classe, onde as primeiras 100 imagens foram adquiridas de forma natural. Outras 900 imagens da base foram criadas por rotações nas 100 imagens descritas anteriormente em 10° , 20° , 30° , 40° , 50° , 60° , 70° , 80° e 90° . A Figura 3 mostra alguns exemplos de imagens por classe contidas na base.

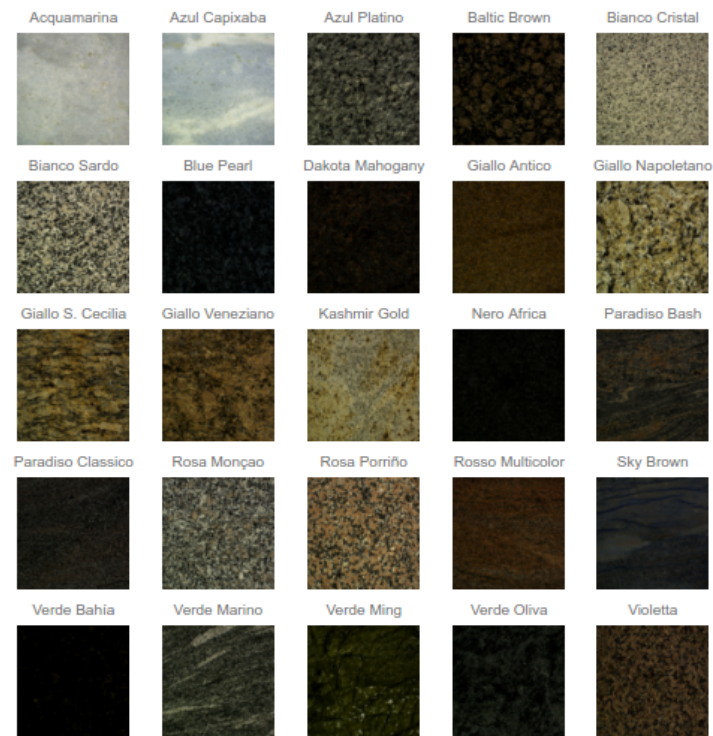


Figura 3 – 25 classes e exemplos da base de de *Bianconi et al.* (BIANCONI et al., 2015) utilizada nos experimentos.

5.1.2 Metodologia Experimental e Métrica Utilizada

Utilizamos em nossos experimentos um protocolo preliminar de validação. Primeiramente separamos 500 das 1000 imagens, contendo 20 imagens por classe, separando-as ainda mais em 250 imagens de treino e 250 imagens de validação, contendo cada subconjunto 10 imagens por classe. Dessa forma podemos encontrar os melhores parâmetros da técnica sem para isso atuar nas 500 imagens de teste restantes, que no âmbito dos experimentos realizados nesse relatório são imagens desconhecidas.

Utilizamos a métrica de acurácia para avaliar a performance de todos os algoritmos, usando a metodologia de validação citada anteriormente. Em um problema multi-classe com c classes, a *Matriz de Confusão* M é construída com c linhas e c colunas. Os c valores

da diagonal principal de M irão mostrar os acertos de cada classe, sendo que valores fora da diagonal principal indicam classificações equivocadas. Assim sendo, a acurácia é calculada como

$$Acurácia = \frac{\sum_{i=1}^c M(i, i)}{\sum_{i=1}^c \sum_{j=1}^c M(i, j)}. \quad (6)$$

onde i e j são índices das linhas e colunas de M respectivamente.

5.1.3 Aspectos de Implementação da Técnica Proposta

Utilizamos em nossa implementação a biblioteca para o MATLAB chamada Mat-ConvNet (VLFEAT, 2016). Estendemos as camadas da rede MNIST1 contida nessa biblioteca, criando as redes MNIST2 e MNIST3, treinando as redes com as imagens de treino, aplicando as imagens novamente nas redes sem a última camada e extraímos os vetores dos blocos por rede individualmente. Logo após, usamos o classificador por *Florestas Aleatórias* nas etapas de treino e validação, procedimento esse detalhado anteriormente na Seção 4. Denominamos as técnicas que usam cada rede individualmente de MNIST1, MNIST2 e MNIST3.

Também concatenamos e normalizamos o vetor de característica (Eq. 4) formado pelos vetores de cada rede, denominando a técnica proposta na pesquisa de *FUSAO_CONCAT*, como descrito na Seção 4.3. Finalmente, testamos a complementaridade das classificações das três redes utilizando votação. Para tal utilizamos cada Rede Neuronal individualmente para extração de características de cada bloco e sua posterior classificação por florestas aleatórias. Dessa maneira, cada par (RNC, classificador) gera um voto de classificação. A classe mais votada entre os três pares define a classe do bloco 28×28 . Denominamos essa variação da técnica proposta de *FUSAO_VOTO*, como descrita na Seção 4.3.

As redes utilizadas foram treinadas em lotes de imagens contendo 100 imagens, com a taxa de aprendizado de 0.001. O método de aprendizado foi por descida de gradiente estocástico com *momentum* igual a 0.9 e queda de peso igual a 0.0005 sem *dropout*. O número de épocas utilizado para o aprendizado de cada rede (MNIST1, MNIST2 e MNIST3) foi encontrado através do menor erro de validação (*valtop1e*) em 18 épocas, e são, respectivamente, 10, 15 e 15 épocas conforme mostra o lado direito das Figuras 4, 5 e 6 a seguir.

Todas as RNCs utilizadas na fusão pela técnica proposta foram treinadas em blocos 28×28 das 250 imagens de treino, sendo que cada imagem 1500×1500 da base original continha 2809 blocos 28×28 . Para amenizar problemas de memória em treinar um classificador usando $2809 \times 250 = 702250$ vetores de característica, optamos por usar apenas 1000 blocos por imagem, treinando o classificador em $1000 \times 250 = 250000$ vetores de característica. O teste ocorre normalmente nos 702250 blocos das 250 imagens de validação. Os experimentos foram executados em um nó de cluster contendo 11 processadores, 131 GB de memória RAM e placa de vídeo NVIDIA GeForce 210.

5.1.4 Estado da arte

Nos experimentos preliminares desta pesquisa utilizamos primeiramente uma série de descritores globais de textura para comparação com a técnica proposta. Aplicamos uma série de descritores conhecidos para aplicações de propósito geral como, por exemplo o *Local Binary Patterns* (denominado LBP nos experimentos) (OJALA; PIETIKÄINEN;

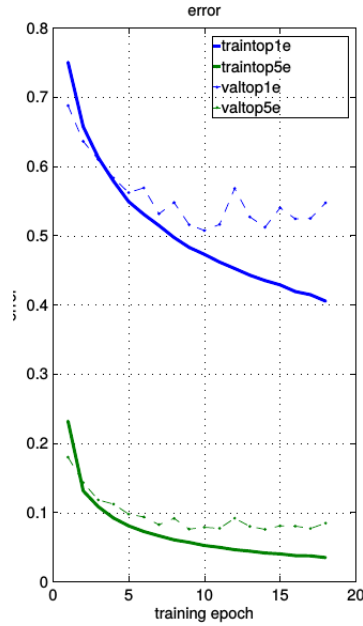


Figura 4 – Resultados de função objetiva e de erro de validação utilizando 18 épocas para o treino da rede MNIST1. Analisando-se a Figura, pode-se perceber que o menor erro de validação (valtop1e) está na décima época.

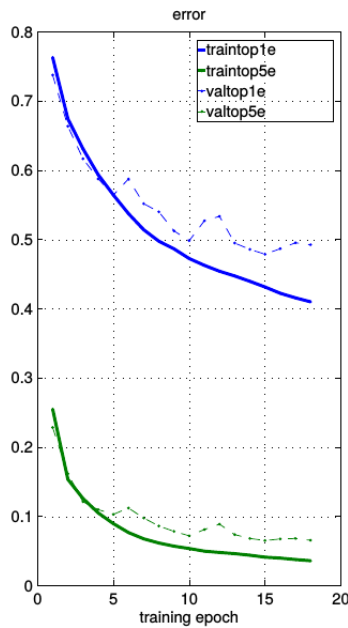


Figura 5 – Resultados de função objetiva e de erro de validação utilizando 18 épocas para o treino da rede MNIST2. Analisando-se a Figura, pode-se perceber que o menor erro de validação (valtop1e) está na décima-quinta época.

HARWOOD, 1996) e o Histograma de Gradientes (DALAL; TRIGGS, 2005) (denominado HOG nos experimentos). O LBP é um histograma de relacionamentos entre o pixel e seus 8 pixel vizinhos. Já o HOG consiste em histogramas de orientações de gradiente em regiões localizadas (retangular ou circular) da imagem.

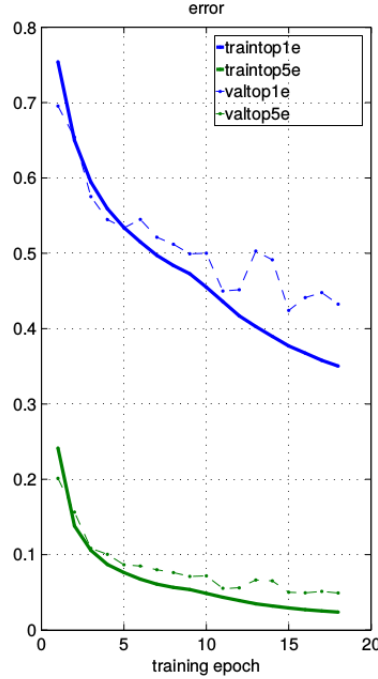


Figura 6 – Resultados de função objetiva e erro de validação utilizando 18 épocas para o treino da rede MNIST3. Analisando-se a Figura, pode-se perceber que o menor erro de validação (valtop1e) está na décima-quinta época.

Outros descritores utilizados nos experimentos foram propostos originalmente para aplicações específicas de caracterização de textura. O Filtro de Gradiente de Textura Convolutacional foi proposto por Ferreira *et al.* (FERREIRA *et al.*, 2015b). Esse descritor caracteriza a textura na forma de histograma em regiões específicas de baixo gradiente logo após a aplicação de filtros multidirecionais de tamanhos variáveis. Tal descritor foi utilizado originalmente para identificação de texturas típicas de impressoras laser. Denominamos as técnicas baseadas em Filtro de Gradiente de Textura Convolutacional nos experimentos para cada tamanho de filtro de $CTGF_{3 \times 3}$, $CTGF_{5 \times 5}$, $CTGF_{7 \times 7}$ e $CTGF_{MS}$.

Os próximos descritores utilizados foram extraídos do mesmo trabalho de Ferreira *et al.* (FERREIRA *et al.*, 2015b), estes são extensões multidirecionais, denominados $GLCM_MD$ e multi-direcionais e multi-escala, denominados $GLCM_MD_MS$, das Matrizes de Co-ocorrência de Nível de Cinza. Finalmente, utilizamos a extensão de Matrizes de Co-ocorrência de Nível de Cinza de Mikkilineni *et al.* (MIKKILINENI *et al.*, 2005), denominado de $GLCM$ nos experimentos. Esta técnica foi utilizada para identificação de padrões de textura de impressoras laser na referência citada.

5.2 Resultados

Utilizando a metodologia e métrica descritos na Seção 5.1.2 na base de dados apresentada na Seção 5.1.1, realizamos os experimentos no conjunto de validação, possuindo a liberdade de variar os parâmetros *número de árvores* do classificador de florestas aleatórias, encontrando dessa forma o melhor parâmetro a ser utilizado em imagens desconhecidas (de teste). Utilizamos valores no conjunto $NA = \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$ para variar o número de árvores das florestas aleatórias. Na Tabela 1 mostramos os resultados dos experimentos em um conjunto de validação. Nos experimentos, aplicam-se as técnicas

do estado da arte descritas na Seção 5.1.4 na imagem inteira e as técnicas propostas (em negrito) classificam a imagem por votação majoritária das classes atribuídas aos seus blocos. O classificador utilizado em todos os experimentos é o de florestas aleatórias.

Método	NA	Acurácia
MNIST3	100	99.6% (249/250)
MNIST2	100	99.6% (249/250)
FUSAO_VOTO	90	99.2% (248/250)
GLCM (MIKKILINENI et al., 2005)	100	99.2% (248/250)
GLCM_MD_MS (FERREIRA et al., 2015b)	60	98.8% (247/250)
GLCM_MD (FERREIRA et al., 2015b)	80	98.8% (247/250)
LBP (OJALA; PIETIKÄINEN; HARWOOD, 1996)	90	98.4% (246/250)
FUSAO_CONCAT	10	98.0% (245/250)
MNIST1	10	97.6% (244/250)
CTGF_MS (FERREIRA et al., 2015b)	30	86.4% (216/250)
CTGF _{3×3} (FERREIRA et al., 2015b)	80	86.0% (215/250)
CTGF _{5×5} (FERREIRA et al., 2015b)	90	85.2% (213/250)
CTGF _{7×7} (FERREIRA et al., 2015b)	90	84.8% (212/250)
HOG (DALAL; TRIGGS, 2005)	100	50.4% (126/250)

Tabela 1 – Comparação de técnicas aplicadas em imagens de nível de cinza, onde NA é o número de árvores utilizado no classificador de florestas aleatórias.

Conforme mostrado na Tabela 1, a técnica proposta MNIST3 consegue uma classificação quase perfeita do conjunto de validação utilizado ao se valer das informações descritivas encontradas na sua última camada, entretanto isso ocorre através da votação dos blocos da imagem. Mostramos na Tabela 2 que a classificação dos blocos dessas redes individualmente é baixa, e que somente a votação por maioria dos votos classifica corretamente as imagens.

NA	MNIST1	MNIST2	MNIST3	FUSAO_VOTO	FUSAO_CONCAT
10	42.27%	49.28%	54.71%	52.59%	53.18%
20	45.38%	51.72%	57.24%	55.10%	56.19%
30	46.86%	52.73%	58.32%	56.09%	57.17%
40	47.86%	53.17%	58.77%	56.55%	57.77%
50	48.44%	53.60%	59.11%	56.88%	58.13%
60	48.78%	53.79%	59.41%	57.04%	58.52%
70	49.16%	53.96%	59.50%	57.21%	58.59%
80	49.40%	54.06%	59.66%	57.35%	58.69%
90	49.63%	54.14%	59.72%	57.42%	58.84%
100	49.72%	54.23%	59.85%	57.47%	58.98%

Tabela 2 – Resultados de classificação dos blocos das redes individualmente (MNIST1, MNIST2 e MNIST3) e das técnicas que envolvem fusão delas (FUSAO_CONCAT e FUSAO_VOTO). NA é o número de árvores utilizado no classificador de florestas aleatórias.

6 Considerações Finais

Nesse primeiro relatório apresentamos uma técnica promissora baseada em Redes Neurais Convolucionais usando a técnica da Multi-Análise, aplicando-a na classificação de imagens de granito. Utilizamos o cenário de *Multi-Dados* da Multi-Análise ao investigar pequenos fragmentos das imagens, ou *patches*, classificando a imagem através da classificação dos seus fragmentos. Utilizamos em conjunto com este cenário investigativo um novo cenário proposto nesta pesquisa, denominado *Múltiplas Configurações*, que é a aplicação de Redes Neurais Convolucionais de diferentes arquiteturas no contexto de *Multi-Dados*.

Mostramos, através de um experimento preliminar de validação, que apesar de as características geradas pelas redes originarem baixa classificação nos fragmentos, o número alto de fragmentos faz com que, por votação majoritária, a classificação das imagens se torne eficiente, como mostrado na Seção 5.2.

Apesar dos resultados iniciais serem promissores, uma série de outros experimentos devem ser realizados de forma a se definir com certeza a eficiência da técnica. Portanto, estabelecemos uma série de trabalhos futuros a serem realizados no restante da pesquisa. Esses trabalhos futuros estão divididos entre os trabalhos de curto e médio prazo, assim como os trabalhos de longo prazo. Os trabalhos de curto e médio prazo para essa pesquisa são:

1. Uso de descritores de cor e de redes Neurais Convolucionais aplicadas em imagens RGB e comparação com as técnicas propostas.
2. Uso de outros descritores de textura e de descritores explicitamente propostos para a classificação de imagens de rocha para comparação com as técnicas propostas.
3. Aplicação do estado da arte em fragmentos das imagens com o uso de votação majoritária para comparação com as técnicas propostas.
4. Aplicação das técnicas nas imagens de teste através do uso de metodologia de validação cruzada 5×2 (DIETTERICH, 1998).
5. Realização de testes estatísticos para avaliar se existe ou não significância estatística nos resultados.

Já os trabalhos de longo prazo para essa pesquisa são:

1. Aplicação de novas técnicas de fusão das Redes Neurais Convolucionais existentes.
2. Criação e investigação da performance de novas Redes Neurais Convolucionais com diferentes arquiteturas.
3. Aplicação de experimentos no âmbito do cenário aberto (SCHEIRER et al., 2013).

Agradecimentos

Os autores agradecem à Profª. Núria Fernández e Bruno Montandon Noronha Barros (CETEM/MCTIC), bem como Millena Basílio da Silva (CETEM-ES/MCTIC) pelas discussões científicas que deram origem a esta linha de pesquisa. Agradecemos também ao PCI-LNCC pela bolsa de pós-doutorado (Nível PCI-DA) que permitiu a elaboração deste trabalho.

Referências

- Araújo Maria et al. Identification of Granite Varieties from Colour Spectrum Data. *Sensors (Basel, Switzerland)*, Molecular Diversity Preservation International (MDPI), v. 10, n. 9, p. 8572–8584, sep 2010. ISSN 1424-8220. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3231240/>>. Citado na página 3.
- BIANCONI, F. et al. On comparing colour spaces from a performance perspective: Application to automated classification of polished natural stones. *Lecture Notes in Computer Science*, n. 9281, p. 71–78, 2015. Citado na página 9.
- BIANCONI, F.; FERNÁNDEZ, A. Granite texture classification with gabor filters. In: *International congress on graphical engineering (INGEGRAF)*. [S.l.: s.n.], 2006. Citado na página 2.
- BIANCONI, F. et al. Automatic classification of granite tiles through colour and texture features. *Expert Syst. Appl.*, Pergamon Press, Inc., Tarrytown, NY, USA, v. 39, n. 12, p. 11212–11218, set. 2012. ISSN 0957-4174. Citado na página 3.
- CARUANA, R.; NICULESCU-MIZIL, A. An empirical comparison of supervised learning algorithms. In: *Proceedings of the 23rd International Conference on Machine Learning*. New York, NY, USA: ACM, 2006. (ICML '06), p. 161–168. ISBN 1-59593-383-2. Disponível em: <<http://doi.acm.org/10.1145/1143844.1143865>>. Citado na página 8.
- DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: SCHMID, C.; SOATTO, S.; TOMASI, C. (Ed.). *Intl. Conference on Computer Vision & Pattern Recognition*. [S.l.: s.n.], 2005. v. 2, p. 886–893. Citado 2 vezes nas páginas 11 e 13.
- DIETTERICH, T. G. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, v. 10, p. 1895–1923, 1998. Citado na página 14.
- FERNÁNDEZ-DELGADO, M. et al. Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.*, JMLR.org, v. 15, n. 1, p. 3133–3181, jan. 2014. ISSN 1532-4435. Disponível em: <<http://dl.acm.org/citation.cfm?id=2627435.2697065>>. Citado na página 8.
- FERREIRA, A. *Multi-Analysis Techniques for Digital Image Forensics*. Tese (Doutorado em Ciência da Computação) — Instituto de Computação da Universidade Estadual de Campinas, Campinas, São Paulo, Brasil, Maio 2016. Citado 3 vezes nas páginas 2, 3 e 4.
- FERREIRA, A. et al. Data-driven feature characterization techniques for laser printer attribution. *IEEE Transactions on Image Forensic and Security (TIFS)*, Submetido. Citado 2 vezes nas páginas 2 e 4.
- FERREIRA, A. et al. Behavior knowledge space-based fusion for copy move forgery detection. *IEEE Transactions on Image Processing*, v. 25, n. 10, p. 4729–4742, Oct 2016. ISSN 1057-7149. Citado 2 vezes nas páginas 2 e 4.
- FERREIRA, A. et al. Laser printer attribution: Exploring new features and beyond. *Forensic Science International*, v. 247, n. 0, p. 105 – 125, 2015. Citado 2 vezes nas páginas 2 e 3.

FERREIRA, A. et al. Laser printer attribution: Exploring new features and beyond. *Forensic Science International*, v. 247, p. 105 – 125, 2015. Citado 2 vezes nas páginas 12 e 13.

FERREIRA, A.; ROCHA, A. A multiscale and multi-perturbation blind forensic technique for median detecting. In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. [S.l.: s.n.], 2014. v. 8827, p. 302–310. Citado 2 vezes nas páginas 2 e 3.

FERREIRA, A.; SANTOS, J. A. dos; ROCHA, A. Multi-directional and multi-scale perturbation approaches for blind forensic median filtering detection. *Journal of Intelligent Data Analysis*, v. 20, n. s1, p. S17–S36, 2016. Citado 2 vezes nas páginas 2 e 3.

Internazionale Marmi e Macchine Carrara S.p.A. *Stone Sector 2010: Italian industry and international trends*. [S.l.], 2011. Citado na página 2.

JOHNSON, J.; KARPATY, A. *CS231n Convolutional Neural Networks for Visual Recognition*. 2016. [Http://cs231n.github.io/transfer-learning/](http://cs231n.github.io/transfer-learning/). Citado na página 6.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Neural Information Processing Systems (NIPS)*. [S.l.: s.n.], 2012. p. 1106–1114. Citado 2 vezes nas páginas 4 e 5.

KURMYSHEV, E. V.; SÁNCHEZ-YÁÑEZ, R. E.; FERNÁNDEZ, A. Colour texture classification for quality control of polished granite tiles. In: *Proceedings of the third IASTED international conference on visualization, imaging and image processing*. [S.l.]: ACTA PRESS, 2013. v. 2, p. 603–608. Citado na página 2.

LECUN, Y. et al. Gradient-based learning applied to document recognition. In: *Proceedings of the IEEE*. [S.l.: s.n.], 1998. p. 2278–2324. Citado na página 4.

LEPISTÖ, L.; KUNTTU, I.; VISA, A. Rock image classification using color features in gabor space. *Journal of Electronic Imaging*, v. 14, n. 4, p. 040503–040503–3, 2005. Citado na página 3.

LIU, M. et al. Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and chinese vinegar. *Sensors and Actuators B: Chemical*, v. 177, p. 970 – 980, 2013. ISSN 0925-4005. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0925400512012671>>. Citado na página 8.

MIKKILINENI, A. K. et al. Printer forensics using svm techniques. In: *Intl. Conference on Digital Printing Technologies*. [S.l.: s.n.], 2005. p. 223–226. Citado 2 vezes nas páginas 12 e 13.

OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, v. 29, p. 51–59, 1996. Citado 2 vezes nas páginas 11 e 13.

SCHEIRER, W. J. et al. Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE Computer Society, Los Alamitos, CA, USA, v. 35, n. 7, p. 1757–1772, 2013. ISSN 0162-8828. Citado na página 14.

VLFEAT. *MatConvNet: CNNs for MATLAB*. 2016. [Http://www.vlfeat.org/matconvnet/](http://www.vlfeat.org/matconvnet/). Citado 2 vezes nas páginas 7 e 10.

YOSINSKI, J. et al. How transferable are features in deep neural networks? In: GHAHRAMANI, Z. et al. (Ed.). *Advances in Neural Information Processing Systems 27*. Curran Associates, Inc., 2014. p. 3320–3328. Disponível em: <<http://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks.pdf>>. Citado na página 6.