



CLASSIFICAÇÃO DE IMAGENS DE TEXTURA NO DOMÍNIO FREQUÊNCIA USANDO REDES NEURAS CONVOLUCIONAIS

GUSTAVO ADOLFO SCHWANTZ OLIVEIRA¹; MATHEUS DA SILVA JAHNKE²;
RICARDO MATSUMURA ARAUJO³

¹Universidade Federal de Pelotas (UFPEL) – gasoliveira@inf.ufpel.edu.br

²Universidade Federal de Pelotas (UFPEL) – mdsjahnke@inf.ufpel.edu.br

³Universidade Federal de Pelotas (UFPEL) – ricardo@inf.ufpel.edu.br

1. INTRODUÇÃO

A classificação de texturas é um processo muito útil no mundo moderno. Ela pode ser utilizada tanto em aplicações de *e-commerce* (classificação de roupas em um site de compras) quanto com objetivos científicos (classificação dos minerais em um determinado terreno). Embora muita vezes seja fácil para um humano classificar uma textura, essa não é uma tarefa trivial para um computador.

Por isso, a classificação de texturas é um ramo de pesquisa em processamento digital de imagens (PDI). É uma das técnicas usadas em PDI para classificar texturas é a transformada de Fourier (FT) (HIPR2, 2019). Basicamente, o que a FT faz é passar uma imagem do domínio espacial para o domínio da frequência. Ou seja, deixamos de enxergar a intensidade dos pixels em função de sua localização para enxergar a intensidade dos pixels em função de sua frequência.

Aprendizado de máquina (ML) é um ramo da inteligência artificial que estuda algoritmos e modelos estatísticos que permitem que computadores realizem uma determinada tarefa sem terem recebido instruções específicas de como fazê-la, baseando-se em padrões e inferência apenas (RASCHKA; MIRJALILI, 2017).

Em ML, uma rede neural convolucional (CNN) é uma classe das redes neurais comumente aplicada na análise de imagens, sendo bastante usada em problemas de classificação. Redes neurais são algoritmos vagamente inspirados nas redes neurais biológicas que constituem o cérebro de animais.

Tais algoritmos “aprendem” a realizar tarefas sendo treinados em exemplos, geralmente sem terem sido programados com uma regra específica para realizar a classificação (GOODFELLOW et al., 2016).

O objetivo deste trabalho é usar uma CNN para classificar a textura em uma imagem. Entretanto, ao invés de receber a imagem no domínio espacial, a rede neural recebe a imagem no domínio frequência. Os autores deste trabalho acreditam que passar a imagem para o domínio frequência pode ajudar a rede neural no processo de classificação da textura.

2. METODOLOGIA

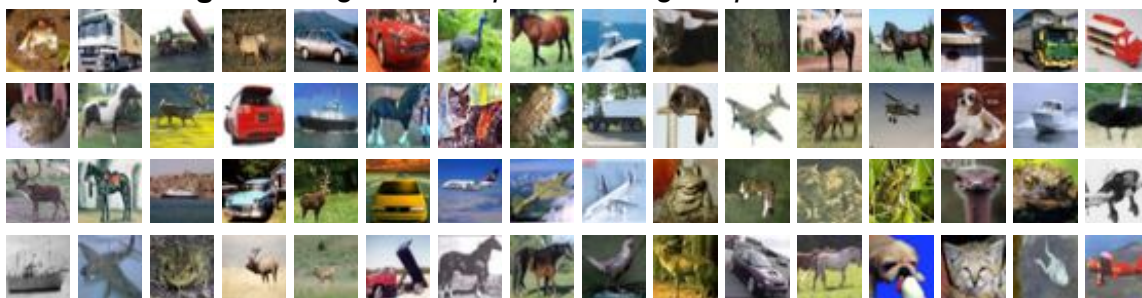
Os datasets utilizados neste trabalho foram o Describable Textures Dataset (DTD) (DTD, 2019), focado em texturas, e o CIFAR-10, que contém imagens de animais e veículos (CIFAR, 2019). O DTD consiste em 5640 imagens divididas em 47 categorias, sendo que cada categoria corresponde a uma textura diferente. Há 120 imagens em cada categoria, e suas dimensões variam de 300x300 a 640x640. Cada imagem do dataset contém pelo menos 90% de sua superfície coberta pela textura em questão. A Figura 1 mostra alguns exemplos de imagens do dataset DTD.

Figura 1. *Alguns exemplos de imagens presentes no DTD cuja texturas são (da esquerda para a direita): sardenta, faixas, espiralada, polvilhada e favo de mel.*



Já o CIFAR-10 é composto de 60000 imagens coloridas divididas em 10 categorias, com 6000 imagens em cada uma. As dimensões das imagens são 32x32. A Figura 2 mostra alguns exemplos de imagens do dataset CIFAR-10.

Figura 2. *Alguns exemplos de imagens presentes no CIFAR-10.*



O projeto usou uma CNN LeNet-5, que é uma rede proposta por Yann LeCun na década de 1990 para reconhecer caracteres escritos à mão (LECUN et al., 1998). O motivo de escolha é o fato da LeNet-5 ser uma rede direta e simples de entender.

Foram considerados três casos distintos: apenas com imagens no domínio espacial, apenas com imagens no domínio frequência e com imagens em ambos os domínios. Nos casos em que as imagens estavam no domínio espacial, foram usadas imagens coloridas (três canais) de tamanho 32x32 e a rede neural possuía 47 saídas (uma para cada textura).



Já nos casos que as imagens estavam no domínio da frequência, elas também eram coloridas de tamanho 32x32 e a rede neural possuía 47 saídas, porém foi utilizado 6 canais nas imagens: 2 canais para cada espectro de fourier dos canais RGB. Também foi feita uma correção gama: $f(x) = x^\gamma$, com γ entre 0 e 1 no domínio da frequência.

Com exceção do último caso, em todas as situações a rede neural usada possuía um bloco convolucional com três camadas convolucionais e duas camadas de max pooling (Conv-MaxPool-Conv-MaxPool-Conv) conectadas a um *multilayer perceptron* (MLP) (120-84-47).

O último caso, que manipulava imagens em ambos os domínios, fez uso de dois blocos convolucionais iguais ao descrito anteriormente, com suas saídas conectadas a uma MLP com entrada de tamanho dobrado (240-84-47).

Em todos os casos, a *loss* foi medida usando-se *cross entropy*. Também, em todos os casos, o otimizador usado foi o *Adaptive Moment Estimation* (Adam), com taxa de aprendizado igual a 0.001. Apenas nos dois primeiros casos (espacial e frequência) foram realizados experimentos usando-se tanto imagens do DTD quanto do CIFAR-10. Optou-se por usar apenas imagens do DTD no terceiro caso (espacial + frequência) pois, como o foco é a classificação de texturas, pareceu desnecessário realizar os mesmo teste usando-se o CIFAR-10.

3. RESULTADOS E DISCUSSÃO

No primeiro caso, que foi o do treinamento usando imagens apenas no domínio espacial e batch de tamanho 32, o maior valor de acurácia ficou em em torno de 90% para o DTD e 50% para o CIFAR-10. Foi constatada uma melhora de acurácia com a diminuição do tamanho do batch para o DTD. Antes do batch ser 32, ele era de tamanho 256 e a acurácia girava em torno de 70%.

No segundo caso, realizou-se o treinamento usando imagens do CIFAR-10 e DTD apenas no domínio frequência. Treinar em cima de imagens do CIFAR-10 no domínio frequência não ajudou muito, sendo obtidas acurácias em torno de 60%. Já ao passar as imagens do DTD para o domínio frequência e usar um tamanho de batch igual a 32, observou-se uma acurácia de aproximadamente 90%. Entretanto, o tempo necessário para treinar foi bem superior ao de quando as imagens deste dataset estavam no domínio espacial.

Por fim, tentou-se usar imagens do DTD em ambos os domínios através de dois blocos convolucionais conectados a uma MLP. Nesse terceiro caso não foi possível concluir o treinamento pois esse excedia o tempo da plataforma Google Colab. No entanto, esse foi o caso com maior acurácia: em torno de 95%. A Tabela 1 a seguir resume os resultados.

Tabela 1. Resultados para diferentes domínios e datasets.

| Domínio | Acurácia | |
|---------------------|----------|-----|
| | CIFAR-10 | DTD |
| espacial | 50% | 90% |
| frequência | 60% | 90% |
| espacial+frequência | - | 95% |

4. CONCLUSÕES

A principal inovação obtida com este trabalho foi poder mostrar que é possível aumentar a acurácia de classificação de uma CNN apenas alterando o domínio dos dados de entrada da rede.

A partir dos resultados obtidos chega-se à conclusão de que passar imagens para o domínio frequência ajuda no processo de classificação de imagens de textura apenas se informações sobre o domínio espacial também estiverem inclusas. Caso contrário, a acurácia permanece a mesma da de quando as imagens estavam no domínio espacial. Também é possível concluir que a passagem para o domínio frequência proporciona um aumento da acurácia para imagens de veículos e animais.

Apesar do processo de classificação melhorar com a passagem para o domínio frequência em ambos os datasets, perde-se consideravelmente em termos de desempenho, com um aumento significativo no tempo necessário para o treinamento se comparado com o caso para o domínio espacial.

REFERÊNCIAS BIBLIOGRÁFICAS

The CIFAR-10. Alex Krizhevsky, Mountain View, mar. 2013-set. 2017. Acessado em 13 set. 2019. Online. Disponível em: <https://www.cs.toronto.edu/~kriz/cifar.html>

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-Based Learning Applied to Document Recognition. **Proceedings of the IEEE**, p. 2278-2324, 1998.

Describable Texture Dataset. Acessado em 13 set. 2019. Online. Disponível em: <http://www.robots.ox.ac.uk/~vgg/data/dtd/index.html>

Fourier Transform. Image Processing Learning Research. Acessado em 13 set. 2019. Online. Disponível em: <http://homepages.inf.ed.ac.uk/rbf/HIPR2/fourier.htm>

RASCHKA, S.; MIRJALILI, V. **Python Machine Learning: Machine Learning and Deep Learning with Python.** USA: Packt Publishing, 2017. 2v.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning.** USA: The MIT Press, 2016. 1v.