

Detecção de um grupo pré-definido de manchas solares com uma rede neural de convolução pré-treinada

Detecting a predefined solar spot group with a pretrained convolutional neural network

RESUMO

Thiago Fellipe Ortiz de Camargo
tcamargo@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil

Marcella Scoczynski Martins
marcella@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil

Virginia Helena Baroncini
virginia@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil

Processamento de imagem e técnicas de aprendizado de máquina são aplicados em diversos campos de pesquisa. Astronomia é uma dessas áreas onde métodos relevantes foram desenvolvidas para identificação e classificação. Este trabalho introduz uma nova abordagem para caracterizar e classificar manchas que aparecem na fotosfera pela expressão de campos magnéticos intensos. Estes campos magnéticos apresentam efeitos significativos na sociedade terrestre. Imagens do Sol foram utilizadas no formato Intensitygram Flat, retiradas do *Helioseismic and Magnetic Imager (HMI)*, do *Solar Dynamics Observatory (SDO)*. Esse método consiste em dois passos: pré-processamento de imagem e uma fase de treinamento usando uma rede neural de convolução (RNC) para identificar e classificar manchas e grupos. Os resultados mostram um grande potencial de processamento que se tornaram um classificador competitivo para os grupos de manchas solares.

PALAVRAS-CHAVE: Processamento de imagem. Rede neural de convolução. Classificação. Manchas Solares.

Recebido: 19 ago. 2019.

Aprovado: 01 out. 2019.

Direito autorial: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



ABSTRACT

IMAGE PROCESSING AND MACHINE LEARNING TECHNIQUES HAVE BEEN APPLIED IN SEVERAL FIELDS OF RESEARCH. ASTRONOMY IS ONE OF THESE AREAS WHERE RELEVANT METHODS HAVE BEEN DEVELOPED TO DEAL MAINLY WITH IDENTIFYING AND CLASSIFICATION. THIS WORK INTRODUCES A NEW APPROACH TO CHARACTERIZE AND CLASSIFY SUNSPOTS WHICH APPEAR IN THE SOLAR PHOTOSPHERE FOR EXPRESSING INTENSE MAGNETIC FIELDS. THESE MAGNETIC FIELDS PRESENT SIGNIFICANT EFFECTS ON EARTH. SOLAR IMAGES WERE CAPTURED IN THE INTENSITYGRAM FLAT FORMAT, FOUND IN THE HELIOSEISMIC AND MAGNETIC IMAGER (HMI), OF THE SOLAR DYNAMICS OBSERVATORY (SDO), WHICH FACILITATE PROCESSING FOR IDENTIFICATION AND SUNSPOTS COUNT. THIS METHOD CONSISTS OF TWO STEPS: IMAGE PRE-PROCESSING AND TRAINING PHASE USING CONVOLUTIONAL NEURAL NETWORK (CNN) TO IDENTIFY AND CHARACTERIZE SPOTS AND GROUPS. RESULTS SHOWED A HIGH POTENTIAL OF THIS PROCESSING THAT BECAME A COMPETITIVE CLASSIFIER FOR THE SUNSPOTS GROUPS.

KEYWORDS: Image processing; convolutional neural network; classification; sunspots.

INTRODUÇÃO

Erupções na superfície solar estão relacionadas com manchas solares e existe a possibilidade de um evento ser estimado se baseado na área, na classe e no ciclo de vida da mancha. Nesta linha, Ejeções de Massa Coronal (EMC) influem significativamente na sociedade da Terra. Entretanto, a predição de *Flares* e seus efeitos ainda são difíceis de serem executados.

Entende-se que o aparecimento de manchas na fotosfera pela expressão de campos magnéticos intensos que bloqueiam a convecção de matéria abaixo da subfotosfera. Essas regiões são mais frias com temperaturas próximas a 4100K, e mais escuras que a subfotosfera. As manchas são formadas por um núcleo, *Umbr*, região mais escura da mancha com dimensões de 300 a 2500km, e com maior força de campo, tendo seu corpo mais distribuído verticalmente. A *Penumbra* aparece em 50% das manchas, com intensidade de campo mais fraca, tendo seu corpo mais distribuído horizontalmente. Atualmente, a referência de medição da atividade de manchas solares é a quantidade de manchas presentes no Sol a qualquer momento determinado

Desde 1981 a análise de imagens provenientes de satélites e observatórios, por processos automatizados, é feito pelo *Solar Influences Data Analysis Centes* (SIDC). Esse centro possui 38 anos e produz mensalmente o *International Relative Sunspot Number, R_i* , que é calculado por através de todos os contribuintes e observações disponíveis, observadores profissionais e astrônomos amadores de vários países, especialmente no leste da Europa, com o *Wolf Number*.

Neste artigo nós propomos a identificação e a classificação de manchas solares através de processamento de imagem para posteriormente ser explorado com intuito de obter a medição da atividade solar. Para este fim, nós utilizamos ferramentas do software *MatLab*, como comunicação numérica e visualização de dados, além de uma rede convolucional pré-treinada (Alexnet). Artigos similares trataram manchas solares em um contexto de processamento de imagem, especialmente utilizando visão computacional, mas, aqui, nossa proposta almeja a aplicação de técnicas de aprendizado de máquina que podem ser exploradas no futuro dentro de propostas de processamento gráfico.

MATERIAL E MÉTODOS

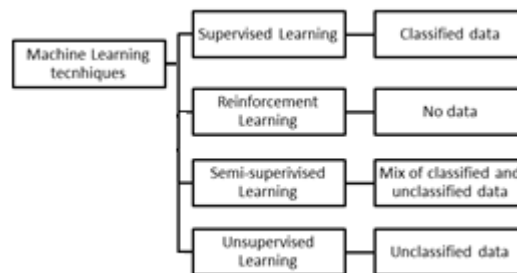
Atualmente, manchas solares são a principal referência para determinar o nível de atividade solar. Além disso, as imagens capturadas do Sol são a principal base para estudos que desenvolvem teorias e para o melhor entendimento da estrela. A qualidade, o bom uso e a facilidade de observação das imagens é essencial e, para aumentar estas características, técnicas de pré-processamento podem ser utilizadas.

Quando se manipula imagens algumas técnicas devem ser aplicadas antes do processo final. Essas técnicas podem ajudar a otimizar os dados: preenchimento de zeros, tratamento de ruído, identificar e remover *outliers* e remover inconsistências; para integrar dados; processamento e redução de dados com valores particulares; normalização e agregação; discretização de dados.

As técnicas de aprendizado de máquina podem ser usados em diversas situações de automatização, pois podem produzir rapidamente e automaticamente modelos capazes de analisar grandes e complexos dados, entregando resultados rápidos e acurados, mesmo em grande escala (Smola 2008 introduction). Aprendizado de máquina é parte de Inteligência Artificial (IA), e apresenta quatro grandes grupos de métodos: aprendizado supervisionado, aprendizado não supervisionado, aprendizado semi-supervisionado e reforço de aprendizado.

As técnicas e suas aplicações estão presentes na figura 1.

Figura 1 – Técnicas de Aprendizado de Máquina



Fonte: Autoria própria (2019).

Este artigo apresenta uma proposta para classificar imagens do Sol de acordo com as manchas e algumas características especiais. Nós endereçamos um banco de dados pego do *Helioseismic and Magnetic Imager* (HMI), que processa imagens do *Solar Dynamic Observatory* (SDO).

Primeiramente, as imagens foram submetidas a técnicas de pré-processamento, como *stretching threshold*, análise de propriedade de objetos e corte, gerando uma subfigura. Um exemplo está presente na figura 2.

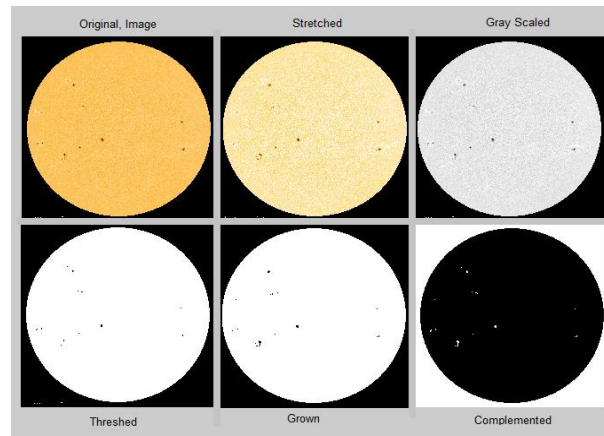
Após o pré-processamento, a rede neural de convolução usa dois grupos de manchas: um com positivo para *O Type* e o outro com negativo. **O O Type é arbitrário**, definido apenas para simplificar a análise, este é caracterizado basicamente por quase apenas penumbra. Uma comparação é mostrada na figura 3. O critério utilizado é baseado na quantidade de pixels pretos em cada subfigura.

Todas as imagens originais foram submetidas a um algoritmo de pré-processamento. A primeira técnica aplicada é o *stretching*, em que o contraste entre as manchas e a superfície solar aumenta. Por seguinte, a escala de cor é alterada de RGB para cinza. O *thresholding* converte em preto a região próxima ao centroide do grupo de manchas, entretanto, em muitos casos o objeto preto criado se mostra dividido em vários centroides pequenos. Por este motivo, o crescimento dos objetos se faz necessário. Esta ação diminui o número de objetos, fato que pode ser considerado como benéfico, entretanto, algumas subfiguras com uma visão diferente do mesmo grupo torna o aprendizado da rede convolucional mais acurada.

A detecção de objetos requer o fundo preto e o objeto branco. Esta condição é limitada pela função *regionprops* do MatLab, assim, o complemento da imagem é obtido, esta função obtém as coordenadas dos centroides dos objetos. Os

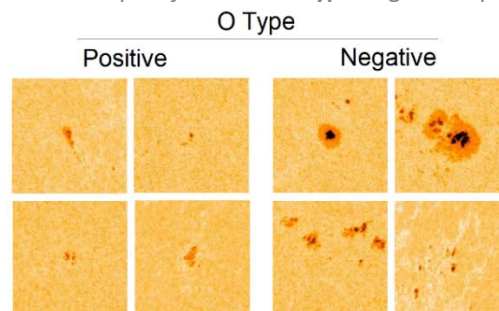
centroides fornecem os parâmetros para o corte da imagem original em subfigura com 227x227 pixels para atender o tamanho de entrada da *AlexNet*.

Figura 2 – Etapas do pré-processamento de imagem



Fonte: Autoria própria (2019).

Figura 3 – Comparação entre **O Type** negativo e positivo



Fonte: Autoria própria (2019).

A última etapa do pré-processamento são dois testes, o primeiro para validar a subfigura, imagens pretas e brancas são descartadas. O segundo teste separa as imagens de acordo com o critério *O Type*, positivo e negativo. Este teste é realizado com base na quantidade total de pixels pretos na subfigura, um limite é determinado, subfiguras que não estão dentro do limite são classificadas como *O Type* negativo.

Um banco de imagens foi criado com duas pastas do grupo *O Type* separados pelo Algoritmo de pré-processamento. O banco de imagem é aleatoriamente dividido em dois vetores, um para treinar a rede, contendo 70% das subfiguras e o outro para validação.

A rede pré-treinada tem suas últimas três camadas modificadas em nosso propósito. As opções de treinamento foram ajustadas visando conservar as configurações originais com um valor de frequência de aprendizado inicial pequeno. Na camada completamente conectada a frequência de aprendizado foi incrementada.

Após o treinamento e classificação, o desempenho da rede foi investigado, calculando-se a precisão, recall e acurácia. Variáveis de controle foram criadas com o intuito de contar a quantidade de positivos verdadeiros, positivos falsos, negativos verdadeiros e negativos falsos. O processo de contagem consiste em

estruturas decisão comparando as rótulo dado pela rede e aqueles do banco de imagem de teste.

RESULTADOS E DISCUSSÃO

No total, 72 imagens no formato HMI Intensitygram Flat Orange 4K foram analisadas, onde 853 subfiguras foram geradas, 253 para testes e 600 para a fase de treinamento. Esse formato permite a fácil manipulação para contagem, identificando e classificando manchas solares do processamento de imagem.

Nós investigamos a performance da RNC calculando três coeficientes precisão, recall e acurácia. O coeficiente de Recall pode ser descrito como a capacidade da rede classificar corretamente, precisão dá a proporção de classificações corretas de um positivo verdadeiro e acurácia determina a fração de classificações corretas.

O quadro 1 apresenta a quantidade das variáveis de controle e valor dos coeficientes.

Quadro 1 – Análise de performance

Positivo Verdadeiro	181
Positivo Falso	18
Negativo Verdadeiro	53
Negativo Falso	3
Recall	0.7391
Precisão	0.9444
Acurácia	0.9170

Fonte: Autoria própria (2019).

A *confusion chart* está presente na Figura 4, onde as linhas representam as classes assumidas, neste caso 0 para tipo o negativo e 1 para positivo. As colunas representam o classe alvo atribuída pela rede neural. A diagonal em verde mostra as classes corretamente classificadas, e a diagonal em vermelho representa as incorretamente classificadas. Por exemplo, 181 imagens foram corretamente classificadas em classe 0, representando 71,5% das 253 imagens testadas. Em outra mão, 18 imagens, ou 7,1% não foram classificadas corretamente para a mesma classe 0. A coluna da direita em cinza representa a precisão destacada em verde, e a taxa de falso descobrimento destacado em vermelho, para cada classe. A linha cinza de baixo representa o Recall destacado em verde e a taxa de Negativo Falso destacada em vermelho para cada classe. A célula azul mostra a exatidão geral destacada em verde.

CONCLUSÃO

Este trabalho investigou um método de identificar e classificar manchas solares com processamento de imagem. Este método consiste em dois passos: pré-processamento e fase de treinamento usando redes neurais convulsionais (RNC).

As imagens endereçadas têm como origem o *Helioseismic and Magnetic Imager* (HMI), do observatório de dinâmica solar. Um total de 72 imagens foram

analisadas, onde 853 subfiguras foram geradas no processo de pré-processamento, separadas em 253 para teste e 600 para o treinamento da RNC

Figura 4 – confusion chart

		Confusion Matrix		
		0	1	2
Output Class	0	181 71.5%	18 7.1%	91.0% 9.0%
	1	3 1.2%	51 20.2%	94.4% 5.6%
	2	99.4% 1.6%	73.9% 26.1%	91.7% 8.3%
		0	1	

Fonte: Autoria própria (2019).

Nós analisamos a performance da RNC de acordo com os componentes precisão, recall, acurácia. A exatidão obtida foi 91.70%, similarmente com alguns trabalhos na literatura. Precisão e recall apresentaram resultados competitivos. Isso significa que a abordagem proposta é um classificador competitivo para os grupos de manchas solares, deixando possível a exploração e extensão para outras imagens relacionadas. No futuro nós esperamos investigar mais técnicas para identificar outras características relevantes do Sol, melhorando pesquisas nesta área espacial.

AGRADECIMENTOS

Thiago Fellipe Ortiz de Camargo agradece o auxílio recebido pela UTFPR para atender a conferencia ColCACCI 2019.

REFERÊNCIAS

GIOVANELLI R. G., **The relations between eruptions and sunspots**. The Astrophysical Journal, vol. 89, p. 555, 1939.

KRIZHEVSKY A., SUTSKEVER I., and HINTON G. E., **Imagenet classification with deep convolutional neural networks**, in Advances in neural information processing systems, 2012, pp. 1097–1105.

SCHWENN R., DAL LAGO A., HUTTUNEN E., and GONZALEZ W. D., **The association of coronal mass ejections with their effects near the earth**, in Annales Geophysicae, vol. 23, no. 3, 2005, pp. 1033–1059.

HOEKSEMA J. T., LIU Y., HAYASHI K., SUN X., SCHOU J., COUIDAT S., NORTON A., BOBRA M., CENTENO R., LEKA K. et al., **The helioseismic and magnetic imager (hmi) vector magnetic field pipeline: overview and performance**, Solar Physics, vol. 289, no. 9, pp. 3483–3530, 2014.