APRENDIZAGEM PROFUNDA APLICADA AO RECONHECIMENTO AUTOMÁTICO DE CARACTERES.

William Bin Falinski (FA), Alceu de Souza Britto Jr. (Orientador), e-mail: alceubritto@gmail.com.

Universidade Estadual de Ponta Grossa/Departamento de Informática.

Ciência da Computação, Metodologia e Técnicas da Computação.

Palavras-chave: CNN, OCR, Redes convolucionais, YOLO.

Resumo

O projeto tem como objetivo analisar e desenvolver soluções de reconhecimento automático de caracteres, ou OCR (Optical Character Recognition), utilizando técnicas de aprendizagem profunda. Para tal serão utilizadas CNNs (Convolutional Neural Network), ou redes neurais convolucionais pré-treinadas para um determinado contexto. Serão avaliados os conceitos de transferência de conhecimento, aprendizagem da representação e a possibilidade de ajuste fino nas redes visando adaptá-las para outros contextos. Espera-se obter como resultados deste projeto de pesquisa e desenvolvimento a geração de modelos neurais para reconhecimento automático de caracteres digitais.

Introdução

A área da visão computacional é responsável por fazer a interpretação do mundo real utilizando quaisquer dados multidimensionais, como imagens e vídeos, que possa representa-lo e traduzi-lo para o mundo computacional. Com o conjunto de dados computados é possível realizar um estudo de abordagem para se aplicar em determinado contexto.

O reconhecimento de padrões e a aprendizagem de máquina são campos diretamente relacionados à visão computacional com o objetivo de realizar a classificação de objetos dentro de um número de classes determinadas. O sistema de reconhecimento pode ser composto de vários módulos, contudo, os pontos cruciais para o bom desempenho se encontram na extração de atributos (representação) e classificação. A escolha de atributos é um passo crítico do projeto e depende de características do domínio do problema.

Em aplicações reais, cerca de 90% do tempo de desenvolvimento é gasto na definição e no projeto de métodos de extração de características. Isso explica o grande esforço que as comunidades de reconhecimento de padrões e aprendizagem de máquina têm dedicado à pesquisa de métodos eficientes de representação.

Recentemente surgiu uma abordagem automática para a extração de atributos. A extração de características e classificação são realizadas pelo mesmo algoritmo, eliminando assim o laborioso processo manual de projetar características. Essa abordagem também é conhecida como representação da aprendizagem. Aqueles que advogam a favor dessa abordagem argumentam que os algoritmos de aprendizado de máquina serão mais eficientes se forem menos dependentes de métodos manuais de extração de características[1].

Uma arquitetura que ganhou bastante popularidade recentemente devido aos excelentes resultados em problemas de classificação foi a Convolutional Neural Network (CNN - Rede Neural Convolucional) [2]. A CNN é um modelo biologicamente

ISSN 1676-0093

inspirado pelos conceitos de campos receptivos e usa em sua arquitetura dois tipos de células, o simples e o complexo. Essas células são usadas para a extração implícita de características de padrões visuais apresentados como entrada e são integrados em uma rede totalmente conectada. Esta rede realiza a classificação usando os recursos extraídos da última camada de células complexas.

Neste projeto pretende-se implementar e avaliar abordagens para reconhecimento de caracteres impressos a partir de imagens digitais utilizando-se redes de convolução, em especial CNNs pré-treinadas. Para tal, utilizaremos conceitos de aprendizagem profunda como transferência de conhecimento e ajuste fino de redes pré-treinadas para adaptar CNNs ao problema de reconhecimento de caracteres digitais impressos. O uso de CNNs pré-treinadas tem como objetivo avaliar o desempenho destas redes.

Material e Métodos

O projeto se encontra dividido nas seguintes etapas:

I. Estudar as soluções existentes de reconhecimento de caracteres em especial as baseadas em aprendizagem profunda;
II. Implementar/instalar técnicas que serão selecionadas para avaliação no projeto;
III. Avaliar solução baseada na extração automática de características usando redes pré-treinadas (transferência de conhecimento);
IV. Avaliar solução usando redes pré-treinadas e adaptadas via ajuste fino;
V. Comparar as soluções geradas em termos de desempenho (taxa acerto) e recursos utilizados.

Analisando o estado da arte das CNNs e suas aplicações, referente a primeira etapa, com os trabalhos desenvolvidos para reconhecimento automático de placas de automóveis[4], classificação de espécies de árvores[3] e identificação e digitalização de uma imagem contendo texto[2], ficam evidentes os resultados promissores que a aprendizagem profunda consegue fornecer para os contextos de visão computacional produzindo resultados com taxas de acerto acima de 90%.

Na implementação, referente a segunda etapa, para a execução da tarefa utilizando transferência de conhecimento e treinamento do zero optou-se pela CNN YOLOv2, a Tiny YOLO[5]. Por ser de pequena dimensão a Tiny YOLO é uma rede que pode ser utilizada em dispositivos que dispõem de capacidade de processamento médio como smartphones e computadores de tamanho reduzido como o Raspberry Pi 3 e o NVIDIA Jetson TX1.

O conjunto de dados escolhidos para o treinamento das CNNs é a seção de caracteres digitais do *dataset* Chars74K que conta com 62 classes contendo 1016 amostras cada em um total de 62992 imagens[6]. As classes estão disponíveis divididas em 10 classes de “0” a “9”, 26 classes de “A” a “Z” e 26 classes de “a” a “z”. Das amostras, 80% foram separadas para treinamento e 20% para validação.

Para o treinamento das redes foi utilizada a aplicação darkflow responsável por fazer a tradução da aplicação original do YOLO, a darknet, para realizar o treinamento através da biblioteca TensorFlow versão 1.5 rodando na linguagem Python versão 3.6. Assim de forma simplificada é possível usufruir da tecnologia cuDNN® e executar o treinamento em uma GPU de forma mais eficiente.

Nas etapas III, IV e V realizou-se uma avaliação comparando a taxa de aprendizagem da rede durante o treinamento assim como a taxa de acerto das duas CNNs nos dados separados para validação. Os resultados principais a serem

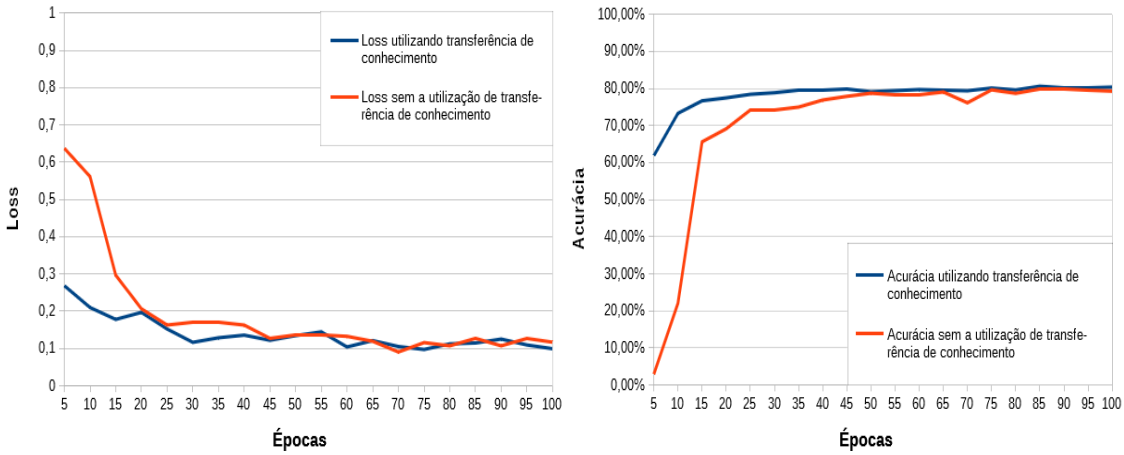
analisados foram a acurácia atingida na base de dados de validação aferindo o quanto a rede neural acerta nas predições das classes da base de dados, e o *Loss* medindo o erro do modelo na base de dados de treinamento.

Resultados e Discussão

Com ambas as redes treinadas utilizando o TensorFlow foi possível realizar as comparações tanto de acurácia como de *Loss.*

A Figura 1, lado esquerdo, mostra os resultados obtidos de *Loss* no decorrer das épocas de treinamento utilizando a base de dados de treinamento. Quanto menor o resultado melhor a rede se sai na predição das imagens utilizando a base de dados de treinamento.

Figura 1 - Gráficos dos resultados da Loss-function (esquerda) e de acurácia (direita) no decorrer das épocas de treinamento.



Fonte: O autor.

A Figura 1, lado direito, apresenta a acurácia dos dois modelos na validação utilizando o conjunto de imagens de validação. A evolução da rede no início do treinamento ocorre de forma gradual e se mantém próximo ao patamar dos 80% a partir da 30ª época, já a rede treinada com os pesos inicializados de forma aleatória, linha laranja, começa com a acurácia baixa e sofre uma melhora abrupta que cessa a partir da 15ª época e se mantém constante próximo aos 80% a partir da 50ª época.

Conclusões

Com os resultados obtidos a partir das redes treinadas com e sem transferência de conhecimento é possível observar as vantagens de se utilizar a técnica, dentre elas a confiabilidade de um extrator de características com um embasamento mais consolidado resultando em valores de acurácia mais altos e economia de processamento trazendo resultados mais rápidos no decorrer das épocas.

A diferença no intervalo da 30ª época para a 50ª época pode não parecer significativa, mas se tratando de gasto computacional, levando se em conta a complexidade da rede e o tamanho do conjunto de dados, a técnica nos oferece melhor desempenho com economia de operações e consequentemente de tempo assim como resultados melhores nas redes.

Na questão relacionada ao conjunto de imagens de treinamento existem problemas como a limitação na quantidade de imagens, onde para se obter bons resultados é necessário uma vasta gama de amostras, e a qualidade das imagens desse conjunto com caracteres legíveis. Uma alternativa para o problema da

quantidade de amostras, se tratando de caracteres digitais, é a ampliação do conjunto que pode ser realizada no computador de forma artificial. Infelizmente para a questão da qualidade das amostras é necessário realizar uma análise ocular, alguns algoritmos conseguem eliminar amostras que estejam corrompidas ou em cores predominantes como no caso do Chars74K o preto ou o branco, mas para o problema de caracteres ilegíveis é necessário uma análise minuciosa.

O contexto da CNN utilizada na transferência de conhecimento está diretamente relacionado com o desempenho da mesma, é preferível utilizar esta técnica para redes que tenham contextos relacionados, como no caso a extração e classificação de caracteres. Um exemplo seria utilizar uma rede pré treinada originalmente para classificar dígitos do alfabeto latino, com a transferência de conhecimento, para utilização em um novo contexto de classificação de caracteres cirílicos.

Agradecimentos

Ao Prof. Doutor Alceu de Souza Britto Jr. pela contribuição com seus conhecimentos e sugestões na orientação deste projeto.

A Universidade Estadual de Ponta Grossa pelos recursos cedidos.

A Fundação Araucária pela oportunidade e por financiar o desenvolvimento do projeto de pesquisa.

A todos que direta ou indiretamente contribuíram para a conclusão deste trabalho.

Referências

1 Bengio, Y.; Courville, A.; Vincent, P. Representation Learning: A Review and New Perspectives. **IEEE Trans. on Pattern Analysis and Machine Intelligence**, [*S. l.*], v. 35, n. 8, p. 1798-1828, mar. 2013. DOI: 10.1109/TPAMI.2013.50. Disponível em: https://arxiv.org/pdf/1206.5538.pdf. Acesso em: 10 mar. 2019.

2 LeCun, Y. *et al*. Gradient-based Learning Applied to Document Recognition. **Proceedings of the IEEE**, [*S. l.*], v. 86, n. 11, p. 2278-2324, dez. 1998. DOI: 10.1109/5.726791. Disponível em: https://www.researchgate.net/publication/2985446. Acesso em: 03 abr. 2019.

3 Haffemann, L.G.; Cavalin, P.; Oliveira, L.S. Forest Species Recognition using Deep
Convolutional Neural Networks. *In*: International Conference on Pattern Recognition, 22., 2014, Stockholm. **Proceedings**[...] Stockholm: IEEE, 2014. p. 1103-1107.

4 Laroca, R. *et al.* A Robust Real-Time Automatic License Plate Recognition Based on the YOLO Detector. *In:* International Joint Conference on Neural Networks, 2018, Rio de Janeiro. **Proceedings**[...] Rio de Janeiro: IEEE, 2018. p. 1-10.

5 Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. *In:* IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2017. Honolulu. **Proceedings**[...]Honolulu: IEEE, 2017. p. 6517-6525.

6 Campos, T. E. de; Babu, B. R.; Varma, M. Character recognition in natural images. *In*: Proceedings of the International Conference on Computer Vision Theory and Applications, 4., 2009, Lisboa. **Proceedings**[...] Lisboa: VISAPP, 2009. p. 273-280.