TÉCNICAS DE APRENDIZAGEM PROFUNDA NA VERIFICAÇÃO DE FACES

Alceu de Souza Britto dos Santos (PIBIC/FA), Ariangelo Hauer Dias (Orientador), e-mail: ariangelo@uepg.br.

Universidade Estadual de Ponta Grossa/Departamento de Informática.

Engenharia elétrica e medidas elétricas, magnéticas e eletrônicas; instrumentação.

Palavras-chave: CNNs, redes convolucionais, verificação de faces.

Resumo

O foco deste projeto foi na avaliação do uso de técnicas de aprendizagem profunda na construção de soluções computacionais de biometria visando a tarefa de verificação faces a partir de imagens digitais. Para tal, foram utilizadas três redes de convolução (CNNs, Siamesas) pré-treinadas para que fosse possível avaliar conceitos como a capacidade de transferência de conhecimento, aprendizagem da representação e a possibilidade de ajuste fino nestas redes visando adaptá-las para diferentes contextos. Esperava-se como resultado deste projeto de pesquisa e desenvolvimento a geração de modelos neurais para a verificação de faces (biometria) em imagens digitais, assim como a produção de conhecimento a ser relatado em artigo científico.

Introdução

Um sistema de reconhecimento de padrões pode ser composto de vários módulos, no entanto, há dois deles que são cruciais para um bom desempenho: extração de atributos (representação) e classificação. A escolha dos atributos é um passo crítico do projeto e depende de características do domínio do problema. Um bom método de extração de atributos fornecerá valores semelhantes a objetos pertencentes à mesma classe e valores diferentes a objetos de outras classes. Se encontrar uma boa representação para um dado problema fosse fácil, a tarefa de classificação seria consideravelmente simplificada. Infelizmente, esse não é o caso da grande maioria dos problemas reais.

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 aprendizado de máquina têm dedicado à pesquisa de métodos eficientes de representação.

Recentemente, com a popularização de algoritmos de aprendizagem profunda, graças ao baixo custo das unidades de processamento gráfico (GPU - *Graphics Processing Unit*), surgiu uma abordagem automática para a extração de atributos. Neste caso, 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, CNN) [2]. A CNN é um modelo

ISSN 1676-0093

biologicamente 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.

Uma rápida revisão da literatura mostra que este tipo de arquitetura produziu melhores resultados do que os métodos clássicos de extração / classificação em vários problemas de classificação, como a classificação de objetos [3], classificação de textura [4] e classificação de caracteres [5]. Mas a CNN tem algumas limitações. Como a quantidade de parâmetros da rede é muito grande, o volume de dados necessários para o aprendizado deve ter a mesma proporção. Em cenários reais, no entanto, a quantidade de dados disponíveis para aprendizagem é limitada.

Neste projeto procurou-se implementar e avaliar soluções computacionais de biometria visando a verificação de faces em imagens digitais utilizando-se redes de convolução, em especial CNNs (Siamesas) pré-treinadas. Para tal, utilizamos conceitos de aprendizagem profunda como transferência de conhecimento e ajuste fino de redes pré-treinadas para adaptar CNNs ao problema de verificação de faces. O uso de CNNs pré-treinadas teve como objetivo avaliar o desempenho destas redes treinadas em bases de imagens diferentes daquelas que foram utilizadas no projeto. Isto na verdade seria uma alternativa para a usual escassez de recursos como bases de imagens e hardware necessários ao treinamento destas redes a partir do zero.

Material e Métodos

Os materiais que utilizamos no decorrer do projeto, foram de extrema importância para a elaboração e desenvolvimento do mesmo, inicialmente utilizamos como apoio ao projeto, a plataforma de hospedagem de código-fonte *GitHub*, a qual nos proporcionou os arquivos e algoritmos necessários para análise e teste, sendo esses postados pelos próprios autores, Davis King, do projeto *ResNet-29*, David Sandberg, do *FaceNet*, e Xy Feng, do *Average Portrait*. Através do *download*, ou clonagem dos códigos-fonte, tivemos acesso à toda codificação de cada projeto.

O próximo passo seria então escolher uma base de dados (imagens), que usaríamos nas avaliações de cada uma das redes convolucionais classificadoras de faces, chegando à escolha da LFW, uma rede vasta de quase 15.000 imagens que nos possibilitou tanto treinar as redes de convolução, quanto classificar posteriormente as imagens.

Escolhidas as redes e a base de dados que utilizaríamos, devíamos escolher agora a *IDE* (*Integrated Development Environment*), que seria o software que seria responsável por compilar os códigos, e receber as devidas edições necessárias nas avaliações das redes. Após alguns testes, foi percebido que os testes seriam muito demorados, pois esse tipo de atividade computacional (utilização de redes convolucionais para treinamento), exige bastante da máquina utilizada.

Sendo assim, optamos pela utilização do *Google Colab,* uma espécie de *IDE* do *Google*, que faz todo o processamento do código, imagem e aprendizagem de forma muito mais rápida e otimizada, que auxilia o desenvolvedor nesse processo de aprendizagem de máquina que demanda muito trabalho de um hardware dito comum. Essa ferramenta otimizou muito o tempo e serviço em cima do projeto, nos proporcionando assim, ainda mais tempo para a fase de testes que dentro desse projeto é de extrema importância.

Foram então efetuados todos os testes necessários, além da análise do código, acurácia, forma de codificação, etapas de processamento e características das redes convolucionais, chegando ao final a ótimos resultados e conclusões que serão expostas mais à frente dentro desse relatório.

Resultados e Discussão

Quanto aos resultados das redes convolucionais, podemos começar falando da rede de Davis King, a ResNet-29, esse modelo que pode também ser reconhecido como *dlib-models*, uma espécie de biblioteca que possui vários tipos de reconhecimento, não só facial, mas também de objetos e outros processos de *machine learning,* possui 29 camadas convolucionais, camadas onde são efetuados os processamentos das imagens e extração das características das fotos, assim, quanto mais camadas, mais detalhada a extração, tendo sempre um número balanceado que nos traz uma acurácia maior.

A ResNet-29 tem acurácia média de 91%, porém esse número pode variar dependendo da base de dados utilizada, a forma de codificação, pois pode haver dentro do código-fonte uma mudança no tamanho da imagem que vem da base. No nosso caso as imagens não sofreram modificações pois vêm com as dimensões todas iguais de 250x250 pixels.

Após a extensa fase de testes dentro da *ResNet-29*, obtivemos ótimos valores de acurácia quando comparado com as médias originais da rede pré-treinada. Nesse experimento chegamos ao valor de 90,4% de acurácia dentro da base LFW.

Falando agora sobre a rede de Xy Feng, a *Average Portrait*, esse modelo é bem interessante pois originalmente trata de um retrato médio de fotos que por ventura sejam inicializadas na entrada do código, porém, um de seus módulos foi adaptado para treinar e reconhecer as imagens da base LFW.

Essa rede possui no decorrer do código a biblioteca dlib, que possui diversos códigos-fonte para diferentes aplicações de *machine learning*, e se diferencia na forma de detecção da face, onde traz todos os contornos dos rostos da base, como evidencia a Figura 1.

Figura 1



Detecção dos contornos da face de uma imagem da base de dados LFW.

Conclusão/ões

A partir de todos esses dados, temos algumas conclusões a tomar, a primeira é que de acordo com os testes, que foram realizados com a mesma base (LFW) para a avaliar as redes convolucionais igualmente, a rede *FaceNet* se saiu melhor quando comparada com a *Average Portrait*, e com a *ResNet-29.*

Desse modo, a rede escolhida para compor a solução monolítica elaborada foi a *FaceNet*, a qual ainda assim gerou ótimo desempenho quando comparada com a rede "pura".

Dito isso, podemos considerar o projeto um sucesso, pois, além da avaliação de três redes convolucionais individualmente, foi também elaborada uma solução monolítica coma rede que obteve um desempenho superior às demais.

Agradecimentos

Agradeço a instituição de financiamento e fomento Fundação Araucária, pelo apoio à esse projeto, também ao meu orientador professor Ariangelo Hauer Dias, o qual sempre sanou todas as dúvidas e esteve presente durante o desenvolvimento do projeto de pesquisa, além da Universidade Estadual de Ponta Grossa pela utilização de suas instalações no decorrer desse projeto.

Referências

[1] Y. Bengio, A. Courville, P. Vincent, Representation Learning: A Review and New Perspectives, IEEE Trans. on Pattern Analysis and Machine Intelligence, 35(8):1798-1828, 2013.

[2] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based Learning Applied to Document Recognition, 86(11):2278-2324, 1998

[3] A. Karpathy et al., Large-scale Video Classification with Convolutional Neural Networks, IEEE Conference on Computer Vision and Pattern Recognition, 1725-1732, 2014.

[4] L. G. Haffemann, P. Cavalin, L. S. Oliveira, Forest Species Recognition using Deep Convolutional Neural Networks. International Conference on Pattern Recognition, 1103-1107, 2014 .

[5] X. X. Niu & C. Y. Suen, A Novel Hybrid CNN-SVM Classifier for Recognizing Handwritten Digits, Pattern Recognition, 35:1318-1325, 2012.