

Relatório 4

Convolutional Neural Networks usando *tensorflow*

1th Pedro Vidal Sales
Universidade Federal da Bahia
Tópicos em Computação Visual 3
Professor: Maurício Pamplona

I. INTRODUÇÃO

Esse relatório explica os parâmetros utilizados para treinar o modelo de rede convolucional para a tarefa de classificar dígitos, utilizando a biblioteca *tensorflow*.

II. DIVISÃO ENTRE TREINO E VALIDAÇÃO

A base de dados utilizada possui 5000 imagens disponíveis para treino, divididas igualmente em 10 classes, uma para cada dígito. A base foi ordenada e foi fixada uma *seed* com valor 1, para que fosse possível recuperar os conjuntos de treino e validação. Após carregar a base, os dados foram permutados aleatoriamente (imagens e labels correspondentes), e depois divididos entre treino e validação. As primeiras 4000 imagens (depois da permutação) foram utilizadas no conjunto de treino, e as outras 1000 imagens foram utilizadas para validação.

III. ARQUITETURA

Todos os modelos possuíam a mesma arquitetura, 3 camadas de convolução, com 32, 64 e 128 filtros, respectivamente e tamanho de kernel 5x5. Cada camada de convolução era seguida por um *max pooling* de *pool_size=2x2* e *strides=2x2*. Após as convoluções e *max pooling*, haviam duas camadas densas, com 128 e 64 nós, respectivamente. A diferença entre cada um dos modelos foi o *augmentation* utilizado.

IV. TREINO

Os modelos A, B, C, D e E foram treinados por 25 épocas, e o modelo F por 200. A cada época (uma passada por todas os exemplos) o conjunto de treino foi permutado aleatoriamente, para que os mini-batches fossem diferentes. Cada mini-batch possui 8 exemplos. O número de passos utilizado foi o número de exemplos do conjunto de treino dividido pelo tamanho do mini-batch, para garantir que cada exemplo da base só seria utilizado uma vez por época. Os valores dos pesos e bias foram atualizados com base no otimizador Adam e na taxa de aprendizado. A taxa de aprendizado que obteve melhores resultados foi 0.0005, ela foi escolhida com base nos trabalhos anteriores. A função de ativação utilizada nas camadas convolucionais e nas camadas *fully connected* foi a função ReLU. A função de ativação utilizada para calcular as probabilidades de cada classe foi a função sigmoid. Os pesos e bias foram inicializados utilizando o inicializador *global_variables_initializer* da própria biblioteca.

V. AUGMENTATION E RESULTADOS

A. Modelo A

Para o treino do modelo A, não foi utilizada nenhuma técnica de *augmentation*, para que o resultado obtido pudesse ser usado para comparação. Este modelo obteve 98% de acurácia.

B. Modelo B

As técnicas utilizadas para modelo B foram translação e rotação. A imagem poderia ser rotacionado em -5.0, -2.5, 2.5 e 5.0 graus, e havia 25% de probabilidade de não haver rotação. Feita a rotação, a imagem poderia ser transladada, em até 3 pixels em qualquer direção. Este modelo obteve acurácia de 98.2%.

C. Modelo C

A única técnica utilizada para o modelo C foi a técnica de translação. Cada imagem poderia ser transladada em até 3 pixels em qualquer direção. Este modelo obteve acurácia de 98.3%.

D. Modelos D, E e F

As técnicas utilizadas para os modelos D, E e F foram rotação, como explicado no modelo B e translação de até 5 pixels. O modelo D difere dos modelos E e F nas probabilidades de sortear os valores de translação, e o modelo F difere do modelo E apenas pelo número de épocas de treino. O modelo E foi treinado por 50 épocas, enquanto o modelo F foi treinado por 200 épocas. Os modelos D, E e F obtiveram 98.3%, 98.4% e 98.9% de acurácia para o *validation*, respectivamente.

E. Ensemble

Foi analisada a acurácia de combinações dos modelos, e o *ensemble* dos modelos B, E e F obteve uma acurácia de 99.1% para o conjunto de validação. Este *ensemble* foi utilizado para colocar labels nas imagens do conjunto *test*, para que pudessem ser utilizadas para treinar novos modelos. Foram treinados novos modelos que utilizavam o conjunto de teste no treino, com e sem as técnicas de *augmentation* citadas anteriormente, mas nenhum destes novos modelos conseguiu ultrapassar a acurácia alcançada pelo *ensemble* dos modelos B, E e F, e nem mesmo alcançaram a acurácia do modelo F, por isso seus resultados foram descartados. Todas as comparações foram realizadas utilizando-se o mesmo conjunto de validação.