

Reconhecimento Facial com a Inception V3

Yago Cruz

UFC - Quixadá

November 26, 2018

1 Arquitetura Inception

- A premissa
- A solução

2 Evolução da arquitetura Inception

- Inception v1
- Inception v2
- Inception v3

3 Conclusões

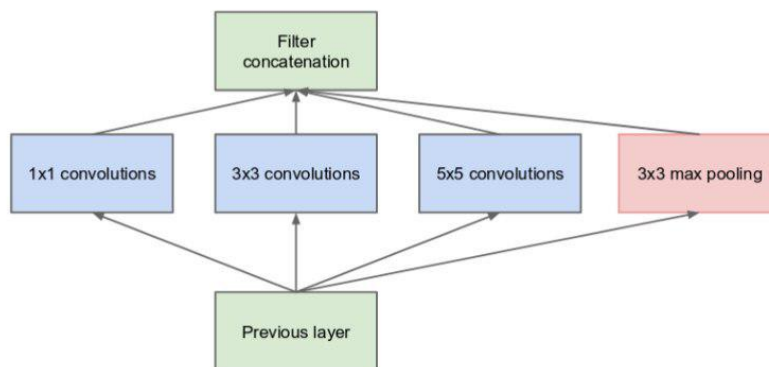
A premissa

- Partes salientes da imagem podem ter variações extremas de tamanho.
- Isso atrapalha a escolha do tamanho correto de kernel. Maior para informações divididas globalmente e menor para informações mais locais.

Arquitetura Inception

A solução

- Utilizar filtros de múltiplos tamanhos para operar no mesmo nível. Fazendo com que a rede fique mais "larga" ao invés de "profunda".
- O conceito de módulos da Inception é realizar convoluções na entrada com 3 filtros diferentes (1x1, 3x3, 5x5). Depois é realizado o max pooling e no final a saída é concatenada e passa para o próximo módulo



(a) Inception module, naïve version

Evolução da arquitetura Inception

Inception v1

- Utiliza de 9 módulos inceptions linearmente alinhados. Vinte e duas camadas e utiliza uma de global average pooling no final do último módulo inception.
- Para prevenir que a metade da rede "morra", foi introduzido dois classificadores auxiliares. Eles aplicam softmax para as saídas dos dois módulos inception e computa uma perda auxiliar sobre a mesma label.

A perda total usada pela Inception

- $total_loss = real_loss + 0.3 * aux_loss_1 + 0.3 * aux_loss_2$

Premissa

- Reduzir o "gargalo representativo". As redes neurais se saem melhor quando convoluções não alteram as dimensões de entrada drasticamente. Reduzir as dimensões podem causar perda de informação, chamado de gargalo representativo.
- Usando métodos inteligentes de fatoração, convoluções podem ser feitas de forma mais eficiente em termos de complexidade computacional.

Solução

- Fatorar a convolução 5×5 para duas de 3×3 para melhorar o desempenho computacional. Mesmo sendo contraintuitivo, uma 5×5 é 2,78 vezes mais cara que uma 3×3 . Fazendo com que colocando duas 3×3 melhore a performance.
- Houve também a fatoração das convoluções dos filtros $n \times n$ para combinações de $1 \times n$ e $n \times 1$. Uma 3×3 é equivalente a realizar uma 1×3 , e em seguida realizar uma 3×1 como saída. Foi descrito que este método é 33% mais barato que uma camada única 3×3 .

Evolução da arquitetura Inception - Inception v3

Premissa

- Foi notado que as funções classificadoras auxiliares não estavam contribuindo muito perto do fim do processo de treino.
- Possibilidades de melhorar a Inception v2 sem alterar muito os módulos.

Evolução da arquitetura Inception - Inception v3

Solução

- A Inception v3 incorporou todas as melhorias da Inception v2.
- Também acrescentou: Um otimizador RMSProp, convoluções 7x7 fatorizadas e Label Smoothing (Um tipo de regularizador de componentes adicionado à função de loss, prevenindo over fitting).

Fim.