

Extração de características de moda utilizando Deep Learning

Vinicius Cin

Departamento de Informática e Estatística - Centro Tecnológico
viniciuscimbr@gmail.com

Objetivos

Desenvolvimento de um sistema completo envolvendo detecção, classificação, segmentação e extração de informações de moda em peças de roupas através de imagens obtidas de desfiles e web scraping. O pipeline tem como finalidade a sua utilização em sistemas de busca automatizada por descrição e similaridade.

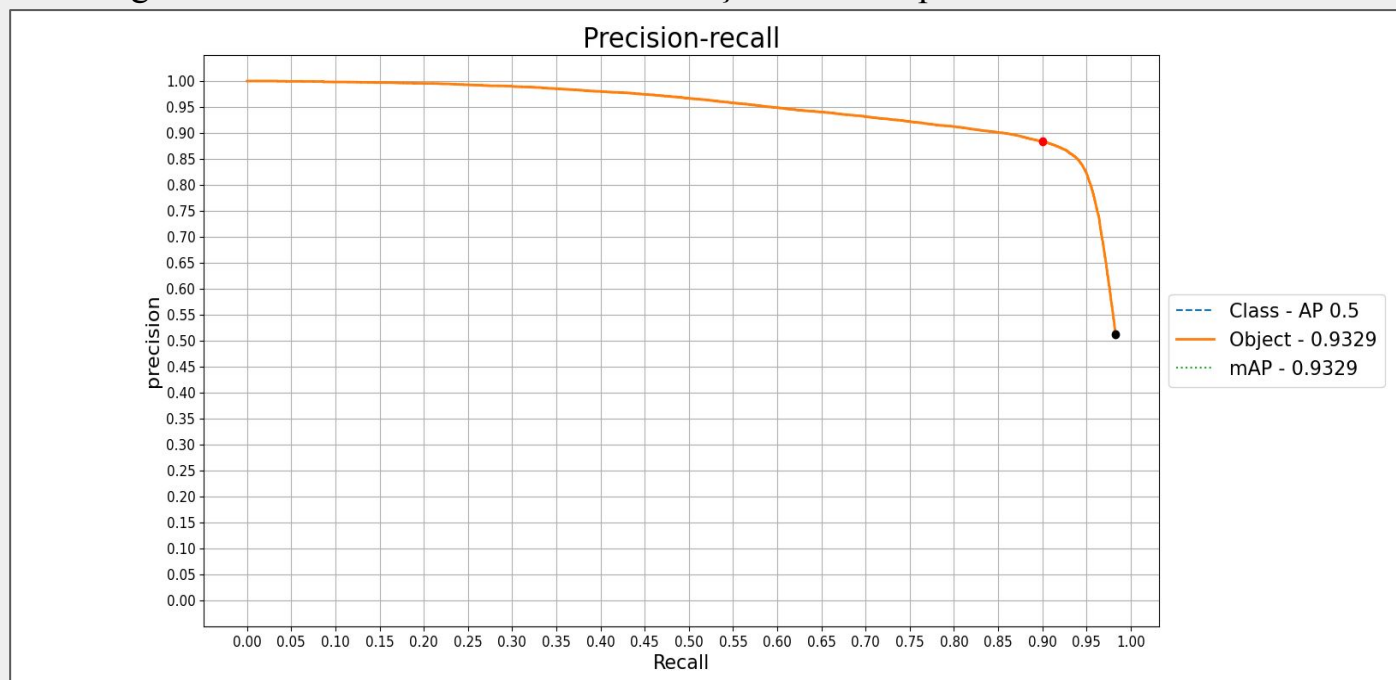
Desenvolvimento

A abordagem inicial consistiu no treinamento de uma rede Yolo v5 como modelo baseline no conjunto de dados DeepFashion, incluindo detecção e classificação das peças. Com isso pode-se observar que devido às peculiaridades do DeepFashion, o modelo não apresentou um recall satisfatório, uma vez que o conjunto de treinamento possui somente uma anotação. O fato do dataset possuir 50 classes também contribuiu para uma baixa precisão, com isso reduziu-se o número de classes de 50 para 16 através da junção de classes similares e exclusão de classes com número de imagens inferior a 2000. De forma a simplificar o problema e obter melhores resultados, foi feita a separação do pipeline de detecção e classificação, passando a utilizar-se uma Resnet para esta tarefa. O modelo de detecção foi então treinado utilizando o conjunto de dados DeepFashion2, que por não possuir as classes de interesse para este trabalho, não foi utilizado na etapa de classificação. Como modelo de segmentação utilizou-se uma rede U2NET pré-treinada no dataset iMaterialist (Fashion) para segmentação de vestimentas em 3 classes: Upper, Lower and Full Body. Similar a UNET, a U2NET usa uma estrutura em U aninhada de dois níveis. Com isso montou-se o pipeline abaixo, que combinado, consegue realizar a detecção das peças, a classificação e segmentação. A extração de características neste trabalho se resume apenas às cores das peças, devido às limitações de tempo.

Resultados

O pipeline é avaliado em duas etapas, primeiramente cada parte é submetida a uma métrica individualmente, desta forma pode-se obter *insights* para o melhoramento individual das partes. Por fim temos a avaliação do pipeline, que tem como objetivo alcançar um recall próximo a 100% com a maior precisão possível.

Figura 1: Curva Precision-Recall - Avaliação de desempenho isolado do detector.



Na detecção obteve-se uma precisão de 51% com um recall de 98% para a classe objeto. Para avaliar o recall nas classes de interesse, utilizou-se uma matriz de confusão personalizada, que nos indica o recall do detector por classe. Obteve-se um recall médio de 95,03% nas classes de interesse.

Figura 2: Recall por classe de interesse - Avaliação de desempenho isolado do detector.



O classificador foi avaliado utilizando-se a matriz de confusão e obtendo-se as métricas de precisão e recall para uma confiança de 25% da classe objeto.

Conjunto de dados

- Classificação: Utilizou-se 16 das 50 classes do dataset DeepFashion e um subconjunto das mais de 290000 imagens disponíveis no dataset, o tamanho médio das imagens é de 300x300, em formato .jpg
- Detecção: Uma vez que o dataset DeepFashion possui somente uma anotação de coordenadas por imagem, utilizou-se o dataset DeepFashion2 para o treinamento de um detector, o dataset possui 191961 imagens para treino e 32153 para validação, com tamanho médio 300x300, formato .jpg divididas em 13 classes.

Figura 3: Pipeline desenvolvido.

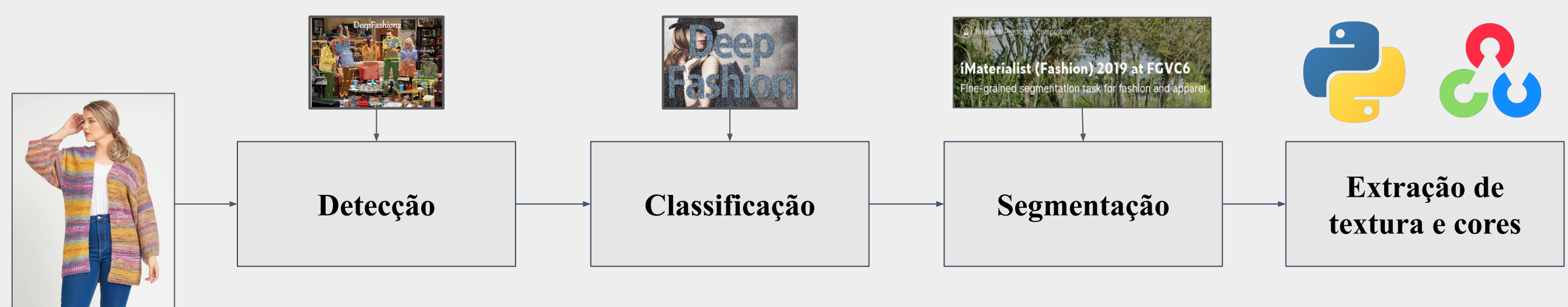


Figura 4: Matriz de confusão
Avaliação de desempenho isolado do classificador.

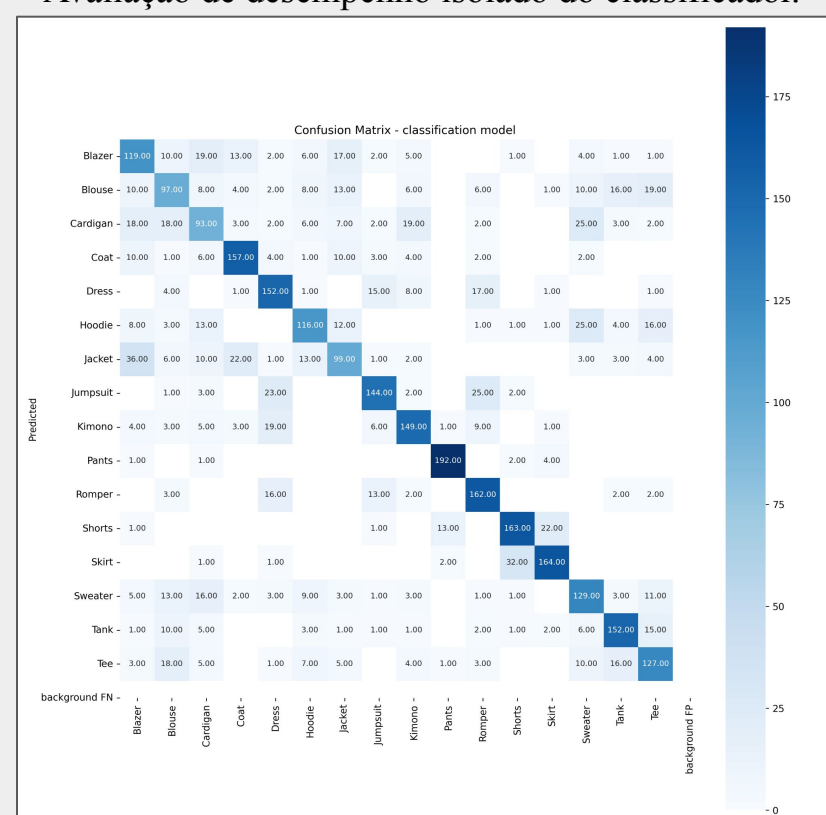
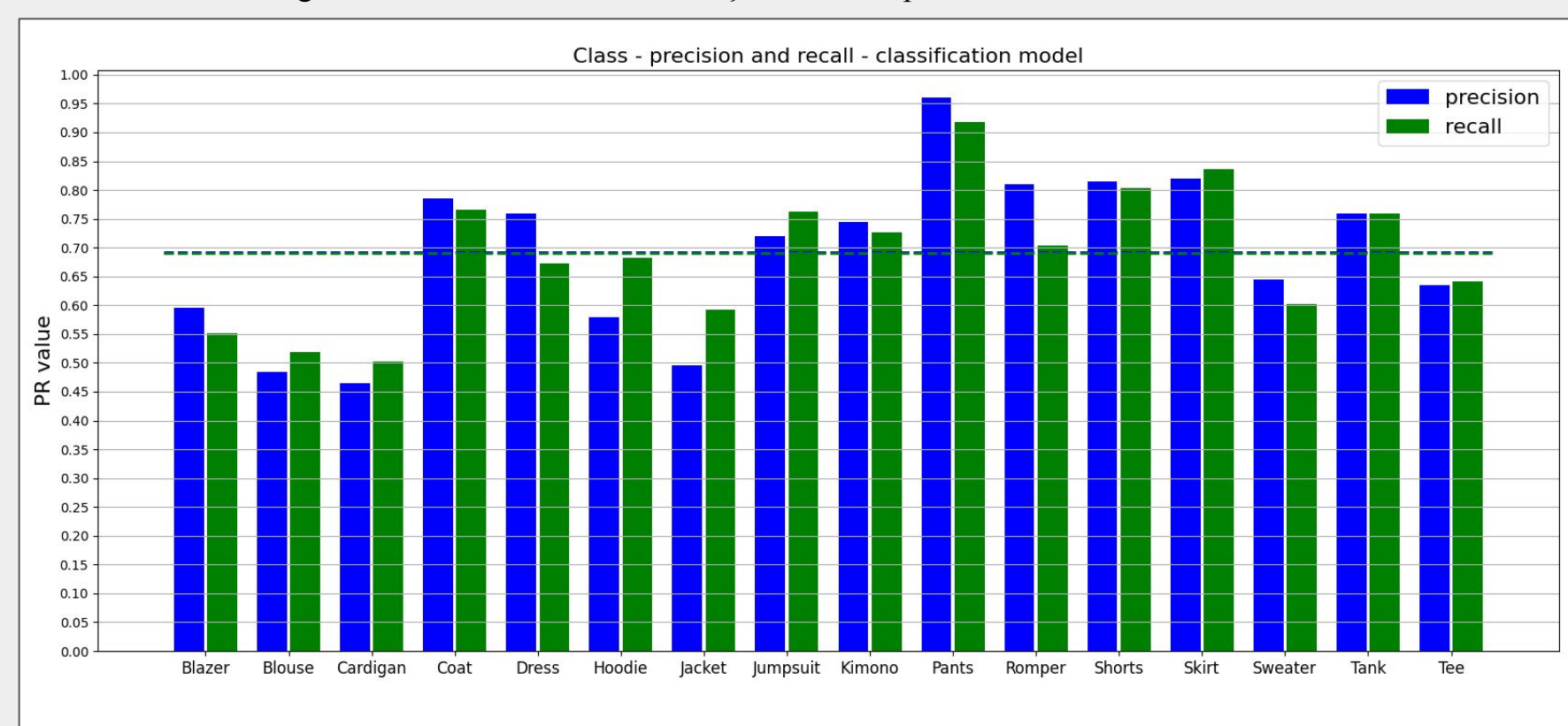


Figura 5: Precisão e recall - Avaliação de desempenho isolado do classificador.



O classificador obteve uma média de aproximadamente 70% de precisão e recall. Ao avaliar o pipeline completo, através da matriz de confusão chegou-se às médias finais de aproximadamente **0.665% de precisão e recall.**

Conclusão

Quando se agrega todas as partes do projeto percebe-se que cada etapa tem efeito significativo na posterior, de forma a limitar o desempenho final, apesar disso, obteve-se resultados satisfatórios para uma primeira versão do pipeline. Com a continuação deste projeto, pretende-se melhorar os resultados e adicionar funcionalidades, como extração de textura, estimativa de tamanho dentre outras.

