

Separação de fotos por Rostos

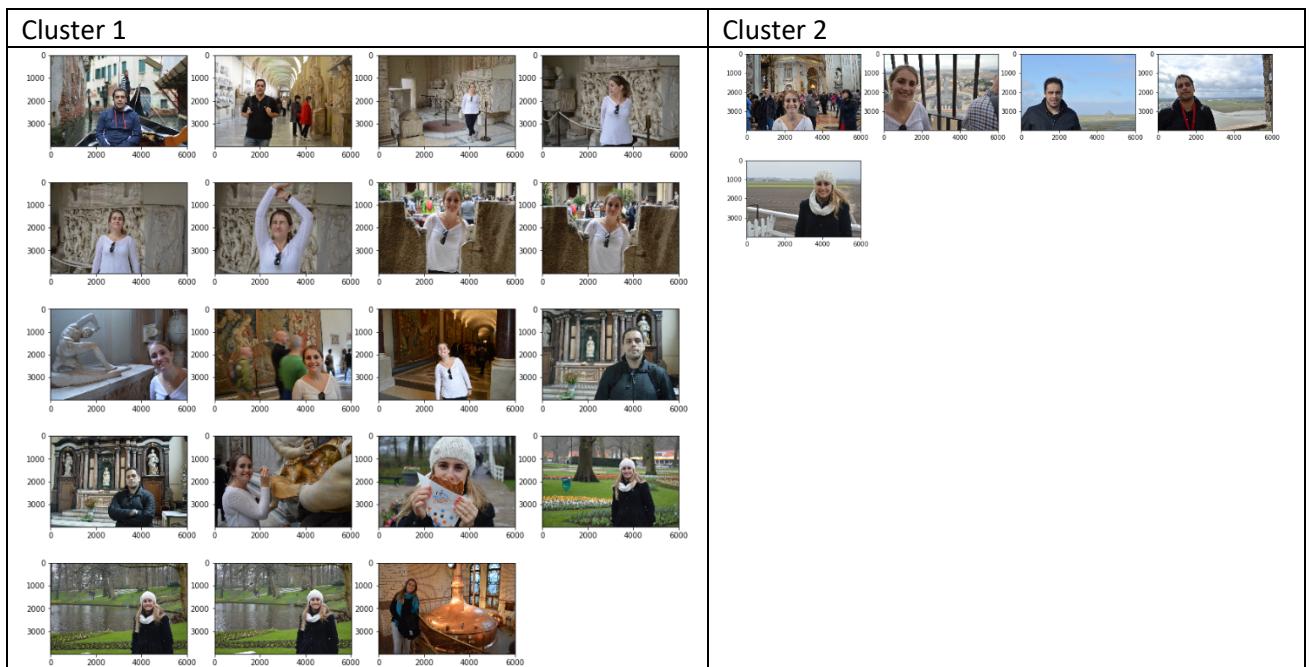
Arquivo: ANS_FOTOS_PESSOAIS.ipynb

Esta seção tem por objetivo verificar de que maneira, utilizando Kmeans, fotos de viagem são classificadas pelo algoritmo. As fotos são de duas pessoas distintas, um homem e uma mulher, que geralmente aparecem em destaque nas fotos, sozinhos.

Para o experimento foi utilizada, inicialmente, a rede **MobileNet** treinada com **imagenet**.

Em um primeiro momento, sem nenhum corte nas fotos foi tentado verificar subjetivamente como as fotos seriam agrupadas.

Com 2 clusters o agrupamento ficou, sem muito sentido, da seguinte forma:



Mais uma avaliação subjetiva foi feita com 4 clusters e, da mesma maneira, nenhum resultado comprehensível pode ser observado. No intuito de buscar novos conhecimentos, as técnicas de validação não serão avaliadas nesta seção como foram exaustivamente mencionadas na seção principal. O objetivo será encontrar maneiras mais criativas de agrupar estas mesmas fotos.

Agrupamento por faces

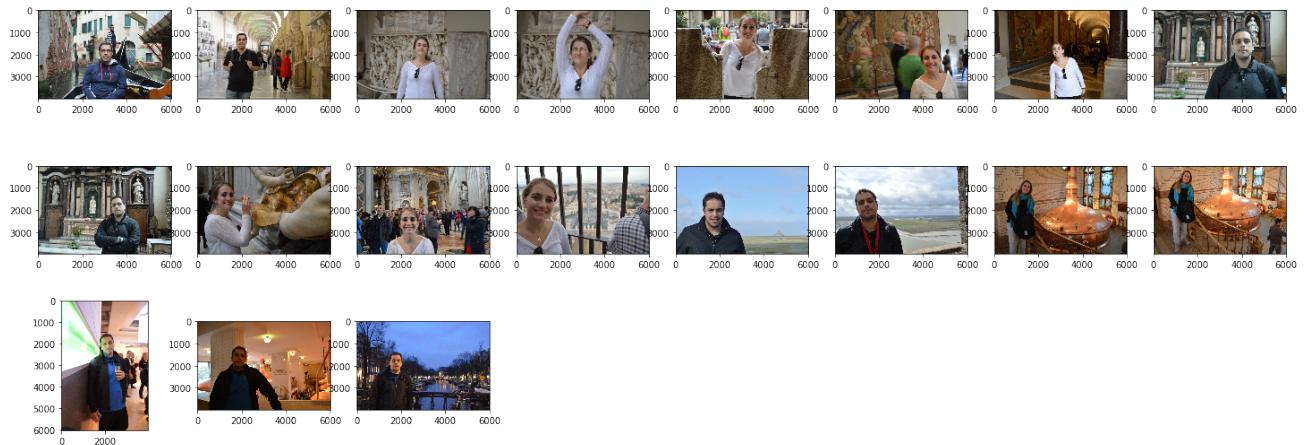
A fim de melhorar o agrupamento das fotos, segmentando os clusters pela pessoa encontrada em destaque na foto, este trabalho contempla duas maneiras distintas de reconhecimento facial. A primeira delas, mais leve porém não tão efetiva, é através da biblioteca OPENCV.

Com esta biblioteca foi necessário ajustar alguns parâmetros e o mais efetivo deles foi o minSize. Como as pessoas a serem segmentadas estão em destaque na foto, a configuração do tamanho mínimo aceito para validar uma parte da imagem como sendo um rosto foi muito importante para que o algoritmo não pegasse rostos aleatórios em segundo plano, estátuas ou objetos sequer semelhantes à uma face.

Porém o algoritmo não conseguiu detectar fotos em 5 imagens, ou seja, estas fotos acabaram sendo ignoradas para fins de agrupamento.

Com K definido com valor dois, os agrupamentos ficaram com 19 fotos no agrupamento um e com 5 fotos no agrupamento 2.

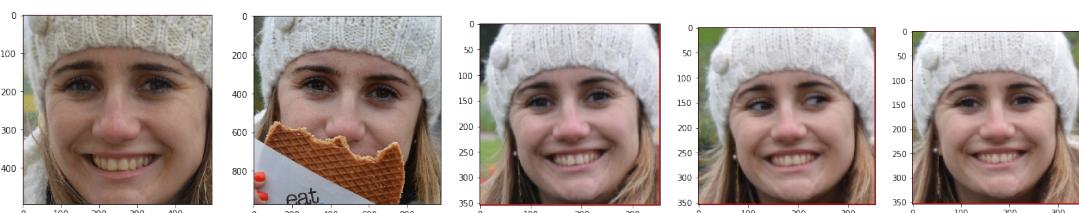
Cluster 1



Cluster 2



Neste momento conseguimos perceber que os rostos acabaram sendo segmentados em função de uma toca utilizada por uma das pessoas, isto é, a rede não segmentou os rostos em função de características faciais, mas sim do adereço utilizado. Isto é compreensível uma vez que o corte realizado nas fotos faz com quem grande parte da imagem estivesse coberta pelo acessório conforme as imagens a seguir:

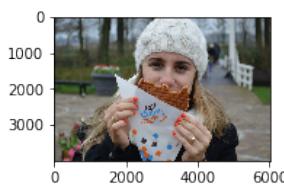


Para tentar compreender melhor como a segmentação está ocorrendo, uma nova tentativa com 4 cluster foi realizada:

CLUSTER 0 - Número de imagens: 16



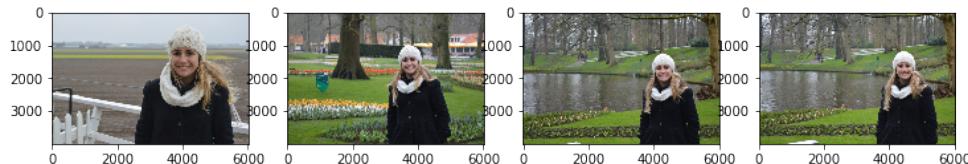
CLUSTER 1 - Número de imagens: 1



CLUSTER 2 - Número de imagens: 3



CLUSTER 3 - Número de imagens: 4

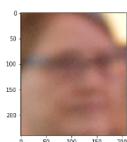


Apesar da existência de inúmeras fotos, uma delas, acabou ficando distante das demais, mesmo ela possuindo o acessório.

A fim de evoluir os resultados, testamos uma rede chamada MTCNN para a extração de faces das fotos. Este algoritmo possui uma precisão impressionante comparada ao método anterior. Mesmo rostos pequenos, distantes da câmera são detectados.



Em virtude dessa característica e como o algoritmo retorna um grau de confiança para cada rosto encontrado, selecionamos então, para esse experimento, apenas as imagens que possuem maior índice de confiança. Infelizmente em duas fotos rostos em segundo plano foram selecionados:



Uma nova tentativa com 4 clusters pareceu mais interessante neste momento.

CLUSTER 0 - Número de imagens: 9



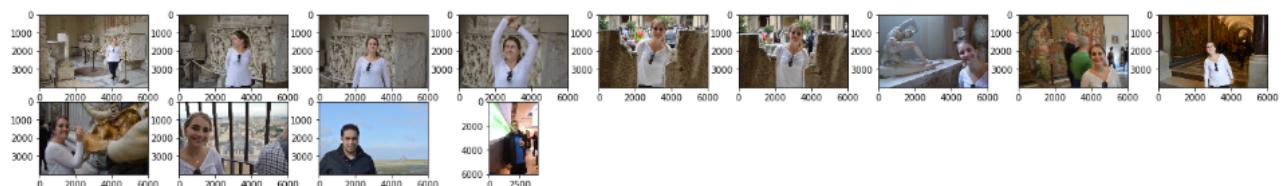
CLUSTER 1 - Número de imagens: 2



CLUSTER 2 - Número de imagens: 5



CLUSTER 3 - Número de imagens: 13



Com exceção de poucas fotos, algumas delas já estão segmentando entre as duas pessoas à serem agrupadas.

Porém ao segmentarmos em apenas dois grupos os resultados não são satisfatórios.

CLUSTER 0 - Número de imagens: 24



CLUSTER 1 - Número de imagens: 5

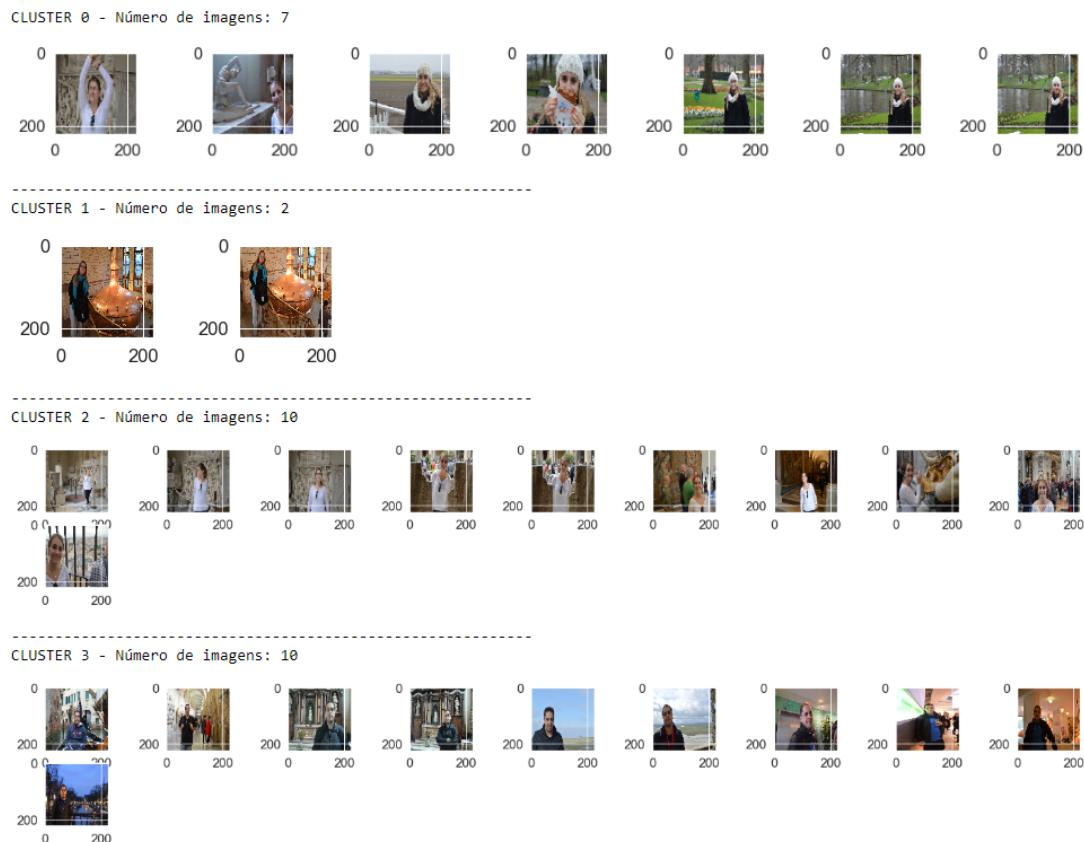


Verificando o índice externo Adjusted Rand Index, mesmo que com um valor baixo, o mesmo maximiza quando selecionados sete clusters. Desta maneira o agrupamento continua apresentando comportamento indesejado.

Desta maneira, uma outra rede, mais indicada para reconhecimento facial foi utilizada. A **VGGFACE**, treinada com uma RESNET50.

O algoritmo de detecção de faces foi alterado em apenas um parâmetro. Antes da MTCNN procurar pelos rostos na imagem, estas foram carregadas para o algoritmo com tamanho reduzido. Desta maneira as faces secundárias que antes possuíam elevado grau de confiança desapareceram e apenas os rostos desejados foram selecionados. De qualquer maneira o desafio continua ocorrendo para aquelas fotos que tiveram os rostos corretamente detectados anteriormente.

Como já havíamos rodado o K-means com 4 clusters repetimos o experimento e os resultados já se mostraram consideravelmente melhores, uma vez que uma das pessoas ficou isolada em um Cluster (3), como pode-se ver na imagem abaixo:



Enfim, o agrupamento mais esperado neste experimento é o de 2 clusters, onde cada “personagem” deveria ficar agrupado sozinho com outras fotos da mesma pessoa. A imagem a seguir mostra como foi o resultado deste teste:

Quantidade de itens em cada classe

[0] 19

[1] 10

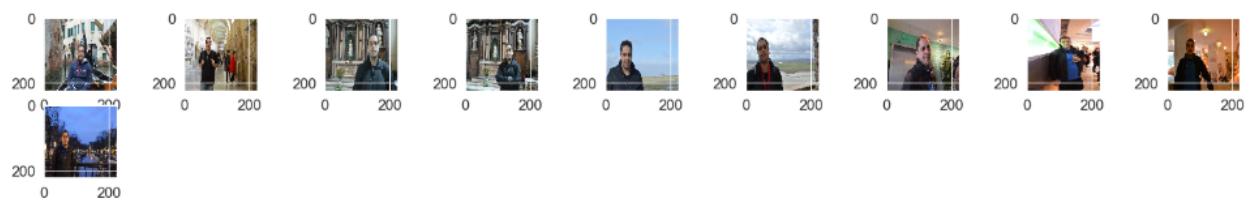
Adjusted Rand Index 1.0

Silhouette Coefficient 0.01

CLUSTER 0 - Número de imagens: 19



CLUSTER 1 - Número de imagens: 10



Como pode-se perceber o agrupamento aconteceu de forma perfeita. Ambos os personagens foram segmentados sem nenhum erro.

Avaliação

Este experimento mostrou que o fato de usarmos modelos corretos ao seu propósito fazem grande diferença no momento de resolver um problema. A Imagenet não foi treinada de forma específica para reconhecimento facial e desta maneira teve dificuldades em resolver o problema proposto de segmentação. Além disto a forma como carregam-se as imagens a fim de que os rostos sejam detectados é de extrema importância para obter-se êxito neste tipo de tarefa. Todas as imagens dos rostos foram recondicionadas ao tamanho 224x224 conforme estas redes foram treinadas, porém apenas o rosto, depois de recortado.