Revisit Cast Searching and Image Retrieval

Chundi Liu
Layer 6 AI

Introduction

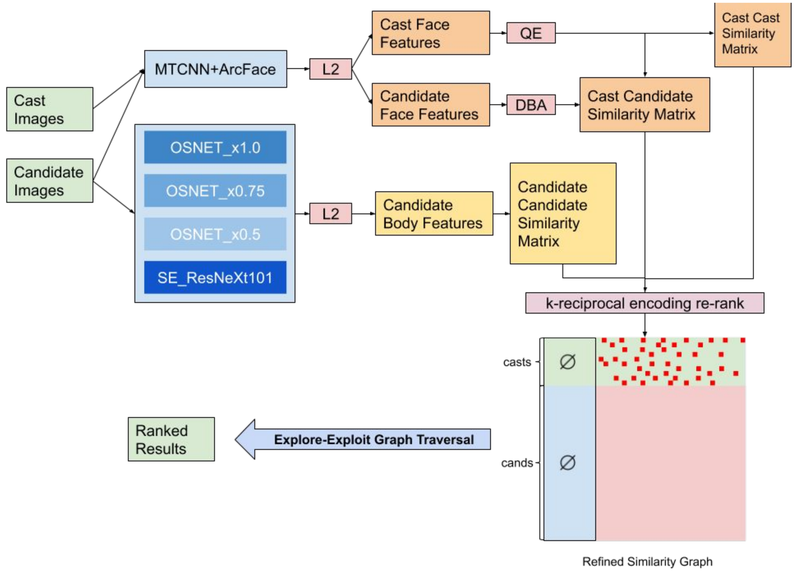
➢ Given an image of a target cast, search for all the candidates (frames of a movie with person bounding boxes) belonging to that cast.

➢ **Challenges:**
➢ Not all candidate images contain frontal face of the cast.
➢ Variations in view angle, lighting and occlusion.
➢ Age variation and makeup.
➢ Camera and person movement.

Solution Details

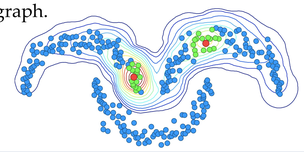
➢ A simple yet efficient two-stage model is applied to solving this problem.
➢ MTCNN is used to learn face detection and fine-tuned ArcFace is used as the face feature extractor.
➢ For generating body features, we trained trained several state-of-the-art person-reid models including Resnet50, OSNET, HACNN and SE_ResNext101 by Cast Search by Portrait 2019 dataset.
➢ Blind training with combination of all training and validation is applied at the end of the competition.
➢ Commonly used k-reciprocal re-ranking is used to further boost the performance.
➢ We also explore many successful techniques in image retrieval (particularly landmark retrieval) area.

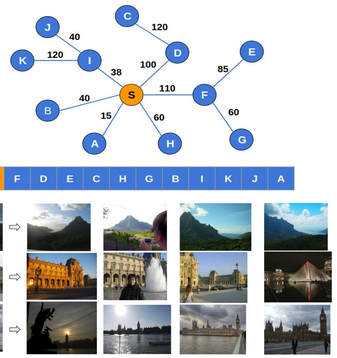
Overall Architecture



Data Manifold and Exploit-Explore Graph Traversal (EGT)

➢ Complex manifold structure commonly exists in real-world data.
➢ Recent work proposes graph traversal based models on k nearest neighbor graph to leverage higher-order neighbor's information
➢ EGT is one of the state-of-the-art efficiently performing retrieval on KNN graph.

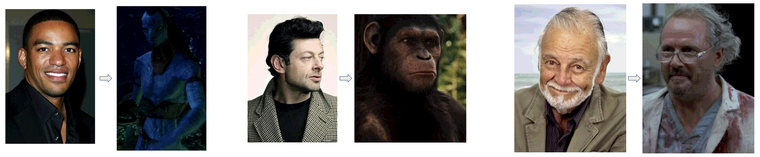




Experiments

Method	Validation	Test
ArcFce	73.34	-
+ Reld Body Feats	79.88	76.86
+QE&DBA	80.51	77.21
+Re-rank	81.76	80.01
+EGT	82.21	80.51
+Blind Training w/ Val Data	82.51	80.58

➢ Failure cases:



Cast Search V.S. Image Retrieval

Cast search problem shares lots of similarity with image retrieval:

- Both of them are ranking problem.
- Both of them are unsupervised problem meaning training on persons (landmarks) from testing dataset is not allowed.
- Data manifold generally exists in both of the problem.

However, we also observe many differences between these two problems. These differences explain why not all the techniques work well on both of the problems.

- In cast search problem, each bounding box contains only one salient person. However, in image retrieval multiple landmarks could exist in one single image and they are equally important, because they are considered as a relevant image to every single landmark.
- People in a movie frame could perform very complicated moves, but the landmarks are rigid. That's why spatial verification techniques which are widely used in image retrieval area is not applicable in cast search problem.
- Movie frames are continuous in temporal space. This provides an unique opportunity for us to capture relevant instances. More likely a person in previous frames would show up in the coming frames.