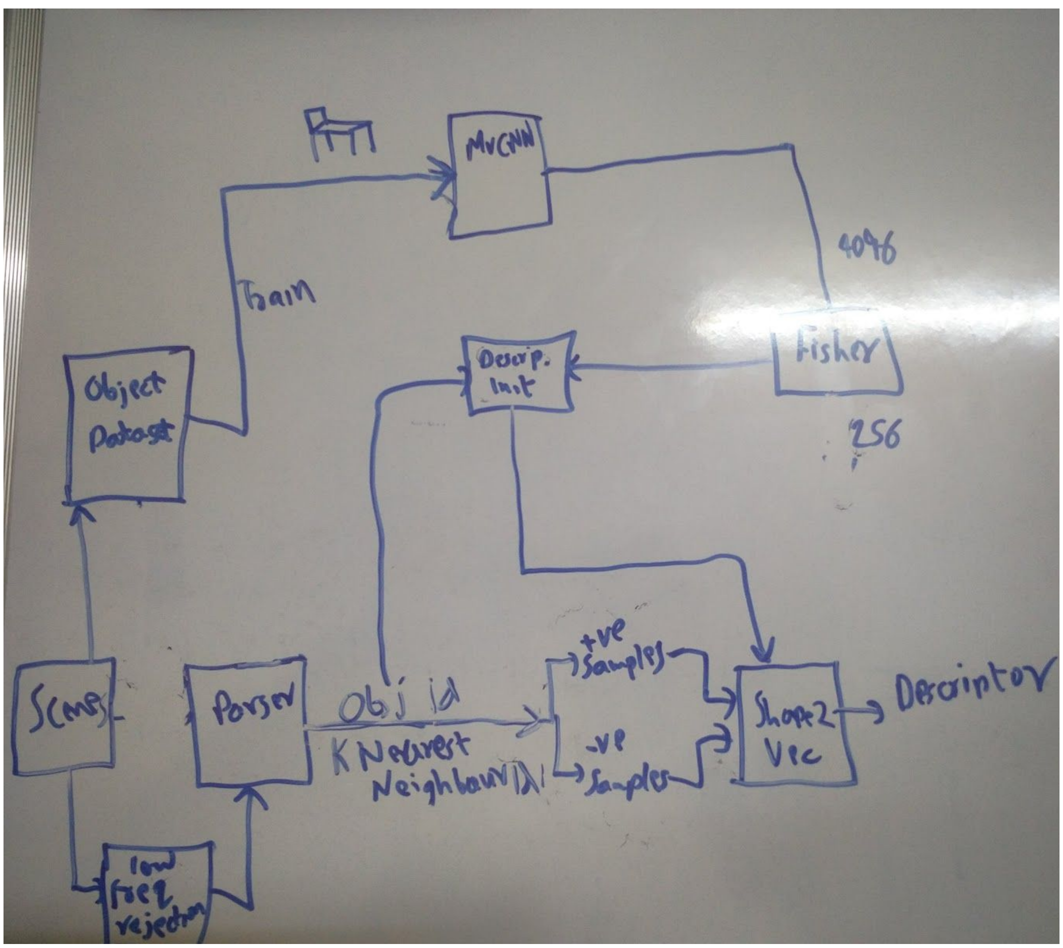
Digital Geometry Processing
Continuous Contextual shape Descriptor
Final Report

Problem statement:
We are trying to find contextual relationship between objects in scenes using continuous representation(shape descriptors) by capturing probable nearby objects from the training data.

Solution:
We are finding shape descriptors which are similar if the objects within the scene are contextually close to each other. It is based on word2vec like embeddings. In word2vec, there is one focus word surrounded by other words which are context word. Word2vec tries to find continuous embedding based on how frequently words co occur together.

Implementation:


Pipeline of the workflow (We used LFD instead of MVCNN in final implementation)

Similar to word2vec, we give each shape a unique id and we randomly initialise their descriptors. In every scene, we find some k nearest neighbours of each object and each close pair as positive sample. Similarly, pairs with objects outside this space form negative samples. Larger the dot product between two descriptors, closer they are. So we treat sigmoid over this dot product to be the probability of objects having similar context. We train over these descriptors to minimise the cross entropy of probability distribution with true labels.

Initialisation:

We initialise the shape descriptors using some state of the art shape descriptor model so as to capture relevant information of structural similarity which might lead to better results and convergence, instead of initialising the descriptors randomly.

We use 2 different initializations for descriptors:

- Light field descriptor + Fisher dimensionality reduction applied over this set
- Angular radial Transform + Fisher dimensionality reduction applied over this set

Results:

We trained our model with a scene dataset with 112 distinct objects, 600 scenes from locations like offices, restaurant, living room, dining room, bedrooms.

Hyperparameters:

Positive:negative ratio 1:10
Descriptor dimensions : 111
Training epochs : 200
Learning rate: 0.2
Optimizer: Gradient Descent with Adaptive moments

Training time GPU ~5mins.

Spot checks**:**

We perform spot checks for objects, we find objects contextually close to the selected objects. We show a few of the results of closeness.

```
mouse
=====================================
mouse                    1.0000
printer                  0.8451
CD_player                0.8344
folder                   0.8083
telephone                0.7623

chair
=====================================
chair                    1.0000
```

```
table                   0.9994
cabinet                 0.9982
sofa                    0.9526
flower                  0.9410
```

```
keyboard
=====================================
keyboard                1.0000
desk                    0.9944
monitor                 0.9895
soundbox                0.9879
desk_lamp               0.9851
```

The results show the the model is able to learn embeddings of nearby objects to be close to each other. In some places due to negative sampling and lack of data, even the negative samples are getting pulled together. E.g. chair and flower are less likely to occur together even then the results are opposite

Analogy test:

Analogy between two pairs of word/shapes means the relation within the pair is same for two pairs. For words, Man:Woman::King:Queen is an example. The analogy for which we tested here is of spatial nearness. So if “Chair is close to Table” and “Mouse is close to Keyboard”, these pairs are analogous.

We try to find similar relationships for shape, eg Chair-table+mouse should give us keyboard. We tried tests like this although most of them are not good, at some places it gives us good analogy.

```
fork:spoon::table:?
=====================================
table 0.828912780799
chair 0.827092709969
cabinet 0.826172446675
sofa 0.772813522925
```

```
fork:spoon::chair:?
=====================================
table 0.828270311323
chair 0.827735179445
cabinet 0.825651011436
sofa 0.771771425576
```

```
fork:spoon::keyboard:?
=====================================
monitor 0.811784875473
```

```
keyboard 0.808962130438
desk 0.808599427064
laptop 0.804642711896
```

Future works:

We fortunately have the labelled data but in generic scenario of the shape segmented scene, we won't have the labels. Thus we can use state of the art shape descriptors to find unique objects in the scene corpus.
Architectural symmetry and structural similarity is also not captured by our model, which can be quite useful in improving the results of contextual closeness.

Instead of taking distance as a binary metric for finding positive samples, a continuous model may be developed which uses this distance value also into account while calculating the contextual closeness values.

We can further work on finding "signed" contextual similarity, which can further capture information like "book on table" as acceptable, as compared to "table on book".

Contribution:

Ashish: Initialisation, dimensionality reduction of the descriptors using Fisher
Aditya: Word2Vec training/testing procedures
Ritish: Word2Vec sampling/preprocessing
Utkarsh : scene preprocessing(object labelling, neighbour detection)