Final Project Progress Report

Team members: Hangjun Piao, Hongyi Wang, Xiaofei Liu, Zirui Tao

The group has a four-phase plan. In the first phase, we would want to achieve simple object detection and segmentation. In the second phase, the program should be able to generate simple text description about the image. In the third phase, the program should be able to generate simple sentences regarding the picture which might be based on templates without contextual information. In the final phase the program would be able to generate a sentence without templates. Currently, we are on the second phase of the plan and working on the data preprocessing.

Image preprocessing : we preprocessed Microsoft COCO image training sets and load the "<imageID,caption>"pair to a redis database. We utilized a pre-trained Alex-net convolutional neural network to extract the features from the last fully connected layer. Each feature vector has a length of 1000. This feature vector would be used as the initial state for the currently building LSTM RNN network.

A snapshot of feature vectors in the database:

idCNNout	CNNoutcol
9	[-1.9006502628326416, 0.3984697461128235, -0.5875108242034912, -2.058480739593506, -0.9624006748199464, 3.049183130264822, 2.6178600788116455, -3.613513708114624, -2....
25	[-1.2976921796798706, -2.2726247310638428, -0.7294253706932068, -0.7335381507873535, -0.9258655309677124, 1.429131865501408, 0.8149363398551941, -2.7783753871917725, -...
30	[-4.505490303039551, 0.49010729789733887, -1.3348559141159058, -2.260587692260742, -0.308530777692794, -0.1853045821189804, -0.9450886249542236, 4.231698036193848, 2....
35	[-0.566621482372839, -1.9155448675155564, -1.7746484565734186, 0.2442614436149597, -2.707285642623902, -2.5815725326538086, -1.283359646797180. 2.1011071205139160 2....
36	[-2.169357061386108. -1.2418845891952517, -2.897099733352661, -3.604495525360107, -2.637493610382008, -0.7338463068008423, -1.777464866381836, -0.7994785159826279, -...
49	[-2.276916265487671, -3.298010349273681, 0.7151650190353394, -0.5125746726989746, -0.942106706695556, -2.385534524917025, -1.671866416931152, 0.250182867050170. -...
61	[-0.1632263362407684, -0.9481610059738416, 0.02904376015067100, -0.9315230846405028, 0.188280671833494566, 0.09792781621217728, 1.413614511489868, 1.255881428718567...
64	[-4.6852941513061520, -4.024137020111084, -2.4246666431427, -2.9239180088043213, -2.055142402648926, -2.099499464035034, -2.6993772983551025, -4.677128791809082, -5.7078...
71	[-3.190721273422241, -4.054117202758789, -0.1926437467336654. -2.6792330741882324, -2.418580532073974, -2.713741540908135, -2.11344075202941. -2.226233482360842. -3...
72	[0.168878734117859, 0.108123816549777798. 0.164766177535057070. 1.731749415397644, 1.644988179206848, 4.946234703063965, 5.253064155578613. 1.224075675010681. 3.1012...
77	[-2.287509202957153. -0.886252582073211. 1.128225326538086, 0.2365385591983795. -0.251384198665189. -3.153282642364502, -1.114517927169998, -2.846571683883667. -3....
78	[-3.010439634323312. 0.5329384803771973. -1.079690098762512. -1.794546489472046. -1.366661190986333. -1.252846598625183. -2.418756008148194. -3.049429416656494. -2.42...
81	[0.716138660907745, -0.997744917869568, 4.335608005523682, 5.9156885147094473, 7.717221260070801, 5.260914325714111, 0.608000814914703. -2.271312236785887. -2.59583...
89	[-3.538757801055908, -1.3998172283172607, -0.7757617235183716, -2.404220115661622, -1.244332075119018. -2.412524461746216. -3.263719081878662. -3.702551364898681. -4....
92	[0.4079079926013946, -1.311544537544250, -0.9097687602043152. 0.456984460353851. 2.424587726593017. 4.762803077697754. 1.6993622779846191. -3.997130155563545. -4...
94	[-3.052208185195923, -3.984418630599976, 0.178827449679374. -2.452099323272705, -1.622542023658752. -2.444789171218872. -4.127129554748535. -2.545598506927490. -3.83...
109	[-1.3378351926803589, -3.091936111450195. 1.727248191833496, 1.0279743671417236. 1.310781836509704. 0.186279773712158. -0.328173518180847. -2.574107885360717. -1...
110	[-1.392669796943664. 0.443875730037689. -0.6229333208924133. -1.923882007598877, 0.4062568843364715. -0.958050012588501. 1.118094325065612. 0.398161590099334. -...
113	[-3.663701534271243, -3.599363803863525. -1.808916091918945. -2.502124071121216. -0.174263626337051. -1.613799452716772. -1.794822692871093. -3.392688512802124. -3.7...
127	[-3.310112842559814. -1.413005232810937. -2.2903492450714113, -1.289922224048. -0.419303743932724. -2.402114868164062. -1.955015182495117. -3.275135278701782. -4.42...
138	[-4.304976463317871, -3.197747073569434, -1.586623787879438. -2.129984617233276. -3.125927448272705. -3.152173280715942. -3.596279621124267. -3.00551056861877. -...
142	[-1.664658069610957, -0.849000334739685, 0.0923912227153778. -0.108500838279974. -0.558061718940735. -1.054875493049621. -0.734069883823394. -3.851811885833744. -...
144	[-3.102679014205932, -0.050802707672119. 0.372616201639175. -2.423392057418823. -2.217324018478393. 1.643947943878173. -0.57619035243980 0.198188289999961855. -...
149	[-0.570027649402618, -1.937938570976257, -0.076554775238037. -1.076557636260986. -1.563477039337158. -0.975874543190002. 0.305073797702789. -0.951326966285705...
151	[-0.421965833478637. -1.165132880210876. 0.0145336724817752. -1.184750437736511. 1.927428364753723. 0.166461288928985. 2.029921770095825. 0.297765672206878...
154	[1.832491517066956, -0.962301135063171. -1.994139671325683. 0.280999514803314. -2.686274290084839. -0.022834293544292. -2.229937553405761. 2.34728097915649...
165	[-2.243971347808838. -3.035481214523315, 0.145945236086845. -0.52773058414459 1.142121434211731. 0.258814321944793. 0.259814321994790. -3.409871578216552. -3.83...
194	[0.976045548915863. 0.966390073299408. -0.778432905673980. -0.445828437805175. -1.214162588119506. 4.601364612579346. 1.599116325378418. -3.430538654327926. -2.658...
201	[-1.234818935394287. -2.143343687057495. 0.4530417621135712. 1.388202190399173. -0.380523473024363. -0.538386694908142. -0.278704047203063. -3.66113615036010. -4.2...
247	[-0.970918953418731. -2.513518571853637. 3.032648324966430. 0.917700529098510. 1.992299437522888. -0.138875424861907. -1.365784168243408. -2.846107244491577. -4....
250	[-2.449009656906128. -2.083570480346679. 0.241232007741928. 0.257676035165786. 0.711570739746093. 1.069855604171753. -1.725049614906311. -1.50935792922973. -2....
260	[-3.067497014999389. 0.767358541486475. -0.371084362684479. 0.174645707011222. -0.261993914842609. -1.067464609146118. -2.057527542114258. -2.167820215225219. -...
263	[0.256268233060836. -3.135310888290405. 0.069634228944778. 0.666555166244506. 1.148892998695373. -0.025927528738975. -0.521659910678863. -1.340067744255066. -...
307	[-0.271279782056080. 0.367585062980651. -1.320982336997985. -3.587958335876465. -1.388267040252685. -1.679962158203125. -0.558899581432342. 2.279513835906982...

Word embedding: the team originally planned to use one-hot representation of the words. However, after learning and discussing with the other groups and the the guest

speaker. The team decided that word vector could be utilized for the project to save training time and increase the efficiency. The team used words from dictionary according to skip-gram method using a dataset cleaned by Matt Mahoney using entire word set from the WikiPedia. The visualized matrix of generated word representation is shown as the following:
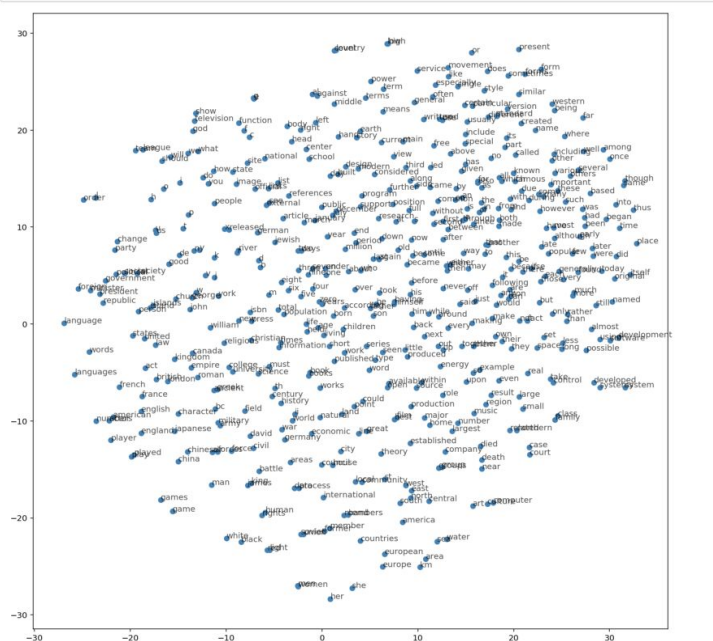
Currently the team is working on designing, implementing and training the LSTM RNN. Although the team has made progress, there are still significant amount of work for the team to do. The team needs to set up the evaluation metric (BLEU), figure out ways to use tensor flow on multiple GPUs, and most importantly, train the network and test the accuracy.

The CPU cycles the team has used: Personal desktops with multiple GPUs on one single CPU.

The library the team used: Raw TensorFlow 1.0 and pretrainned networks.