UC San Diego
JACOBS SCHOOL OF ENGINEERING

Ruoqi Zhang ruz005@ucsd.edu
Kai Wang k7wang@ucsd.edu
Yuekuan Luo y8luo@ucsd.edu
Tianran Zhang tiz217@ucsd.edu

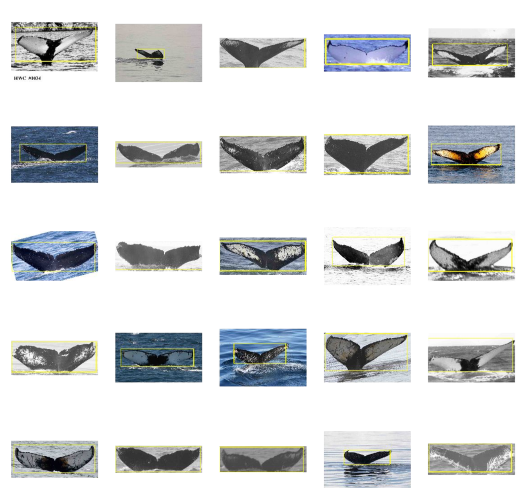
18 The Identification Of The Endangered Whale

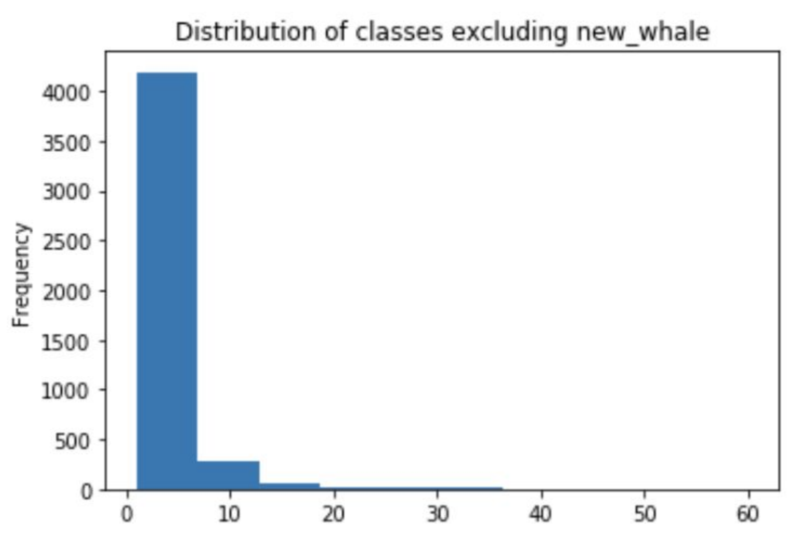
Predicting

We are trying to use machine learning to identify the species of whales by analyzing their' tails and unique markings found in footage. We downloaded the data from a Kaggle competition and separated the data into two parts: 80% as training data and 20% as test data. We used CNN to extract 64 features from the train dataset and then implemented five models to compare their performances.

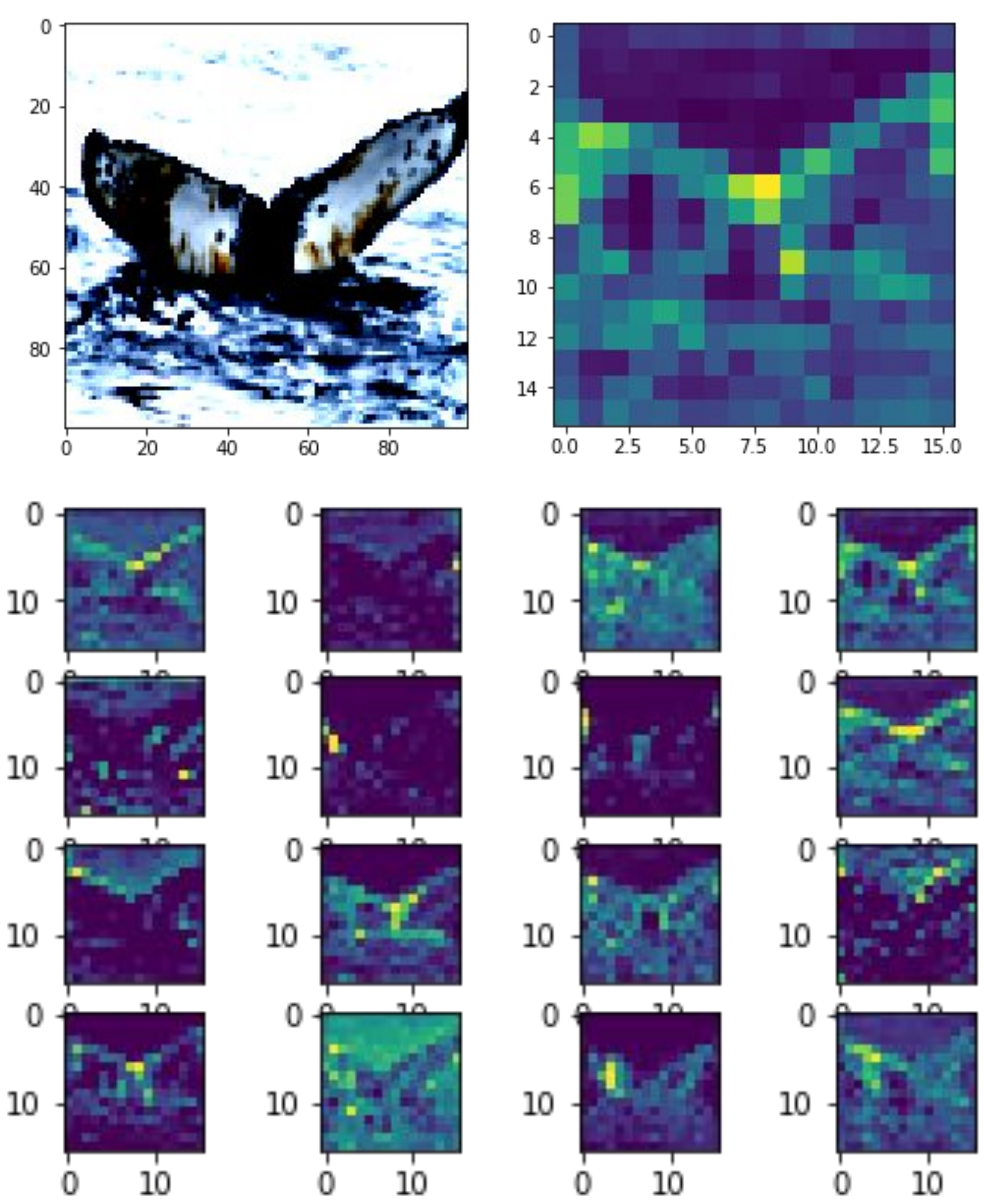
Data

There are 25361 images in the dataset with 4571 unique classes (Whale ID). The largest class is *new_whale*, which is the label of unrecorded humpback whale. And most classes(more than 4000) contain less than 10 images.





Features



Here we use the CNN network to extract higher level features from the image. There are totally 64 channels of features part of which are visualized above.

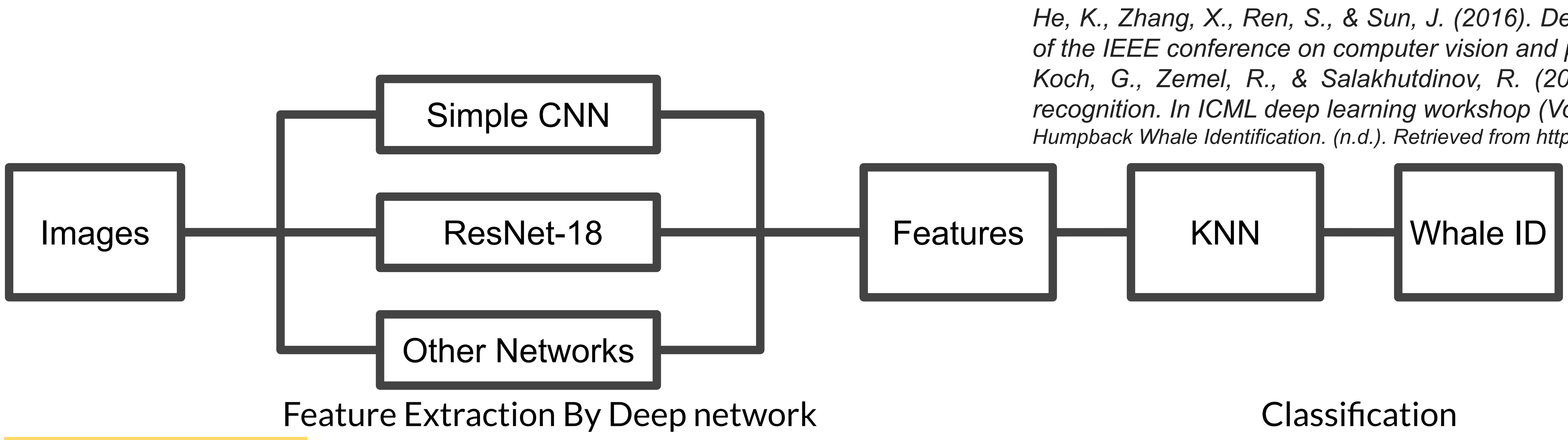
Models

In this project, we totally utilized five models:

- Simple CNN Network
- Simple KNN
- ResNet-18
- ResNet-101 (Transfer Learning)
- Ensemble Network Inspired by Siamese Network

We utilized simple CNN, ResNet-18, ResNet-101 and Ensemble Network to extract features, then use KNN to judge the species of the Whales.

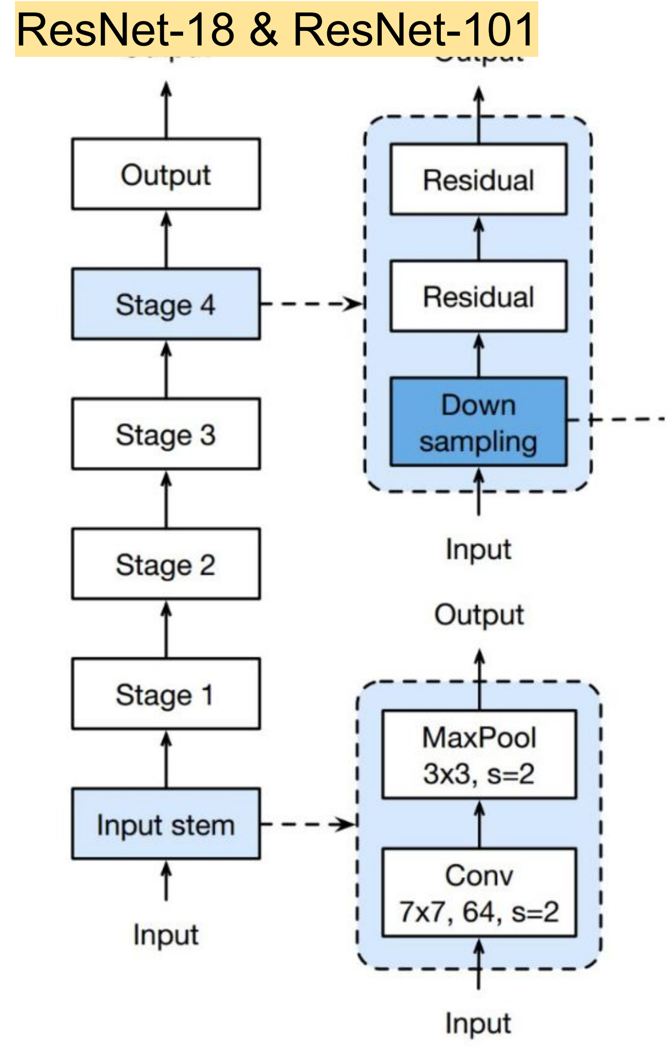
Ensemble Network inspired by Siamese Network



Simple CNN Network

Essentially neural networks that use convolution in place of general matrix multiplications at least for the first layers (Just like the following figure). They are designed to process the data in the form of multidimensional arrays/tensors and always have three main gredients: Local receptive fields, Shared weights and Pooling

ResNet-18 & ResNet-101



ResNet Network is a kind of advanced CNN. This network tries to fix the degradation problem. ResNet Network Each network consists of three main parts: the input part, the output part, and the intermediate convolution part (the intermediate convolution part includes a total of four stages from Stage1 to Stage4 as shown). Although there are many variants of ResNet, they all follow the above structural features. The difference between the networks is mainly due to the difference in the block parameters and the number of intermediate convolutions.

Siamese Network

It's a merged network with two or more subnetworks. Those networks are identical and they share the weights, just like siamese twins. By extracting features, the connected part will compare the L-2 distance to classify the images.

KNN

An algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by a distance function. If K = 1, then the case is simply assigned to the class of its nearest neighbor.

Distance Function

Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$ Manhattan $\sum_{i=1}^{k}|x_i - y_i|$

Ensemble Network

The best performance we implemented is ensemble network. Inspired by siamese network, we ensembled the cnn model and knn model. The pre-trained cnn network extract the features from augmented data, which is classified by knn then. This model fixed the low efficient problem of original siamese network. Knn model is perfectly fit the L-2 distance classification problem.

Results

Models	Test Accuracies
Simple CNN Network	0.32
Simple KNN	0.486
ResNet-18	0.37
ResNet-101 (Transfer Learning)	0.38
Ensemble Network Inspired by Siamese Network	0.4764

Train Set: 20288 samples **Test Set**: 2073 samples

Model Training time:
Simple CNN 10 min per epoch
Resnet-18 20 min per epoch
Resnet-101 45 min per epoch
KNN 40 - 120 min totally

Discussion

Comparing the results, CNNs have the limited ability to classify the images with such large number of classes. However, CNN has a strong feature extraction ability which can be utilized before classification. The combination of feature extraction network and L-2 distance classifier has a better performance than a single simple model.

Future

Due to the complexity of this project, the results are still not good enough with a high price to learn and predict. Another problem is KNN has the limitation about the memory. Also due to the large scale of data, the memory size has a really bad influence on the efficiency of KNN model, which does not match our expectation to improve the problem of Siamese Network.
Our next step includes: 1) Try some faster models, i.e. DenseNet + GBDT; 2) Implement some data preprocessing tools, i.e. *Class merge*, *Bounding box*.

References

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
Koch, G., Zemel, R., & Salakhutdinov, R. (2015, July). Siamese neural networks for one-shot image recognition. In ICML deep learning workshop (Vol. 2).
Humpback Whale Identification. (n.d.). Retrieved from https://www.kaggle.com/c/humpback-whale-identification