Reducing Risks in Open Set Recognition

Xueyang Yu Yijie Fan Xinchen Jin

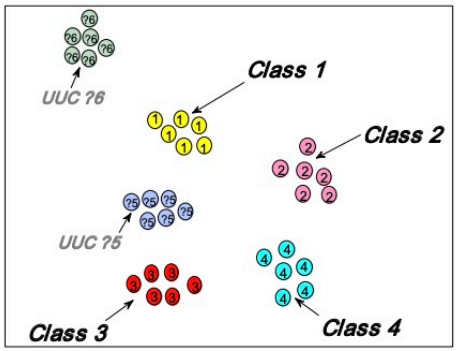
School of Information Science and Technology
ShanghaiTech University

Jun. 20, 2022

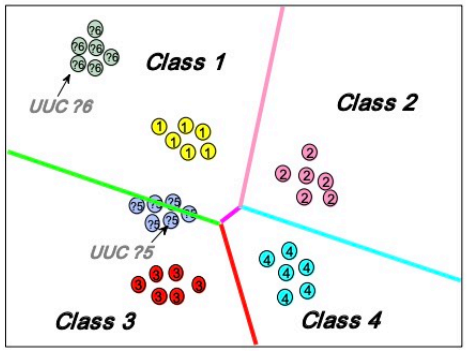
Summary

Motivation

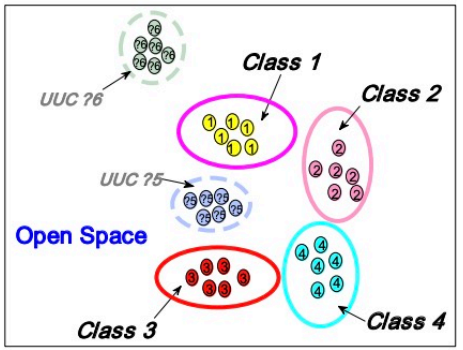
Why is open set recognition important



(a) Distribution of the original data set.



(b) Traditional recognition/classification problem.



(c) Open set recognition/classification problem.

Traditional classification algorithm

Data distributed in closed sets

Open set recognition

Out of distribution data may appear

What are the benefits to reject unknown

Figure: Face Recognition and Autonomous Driving

Problem Definition

Open Set Recognition

First, consider a labelled training set for a classifier:

$$\mathcal{D}_{\text{train}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N} \subset \mathcal{X} \times \mathcal{C}.$$

Here, $\mathcal{X}$ is the input space and $\mathcal{C}$ is the set of 'known' classes. In the closed-set scenario, the model is evaluated on a testing set in which the labels are also drawn from the same set of classes, *ie*,

$$\mathcal{D}_{\text{test-closed}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{M} \subset \mathcal{X} \times \mathcal{C}.$$

In the closed-set setting, the model returns a distribution over the known classes as $p(y|\mathbf{x})$.

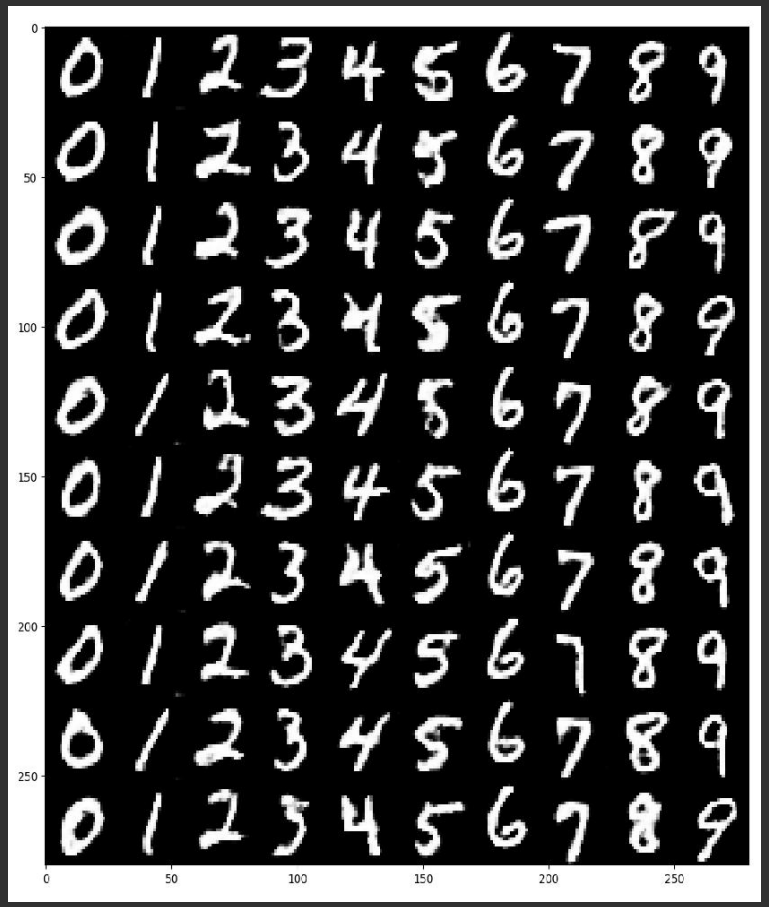
Conversely, in OSR, test images may also come from unseen classes $\mathcal{U}$, giving

$$\mathcal{D}_{\text{test-open}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{M'} \subset \mathcal{X} \times (\mathcal{C} \cup \mathcal{U}).$$

In the open-set setting, in addition to returning the distribution $p(y|\mathbf{x}, y \in \mathcal{C})$ over known classes, the model also returns a score $\mathcal{S}(y \in \mathcal{C}|\mathbf{x})$ to indicate whether or not the test sample belongs to *any* of the known classes.

Data generation and result evaluation

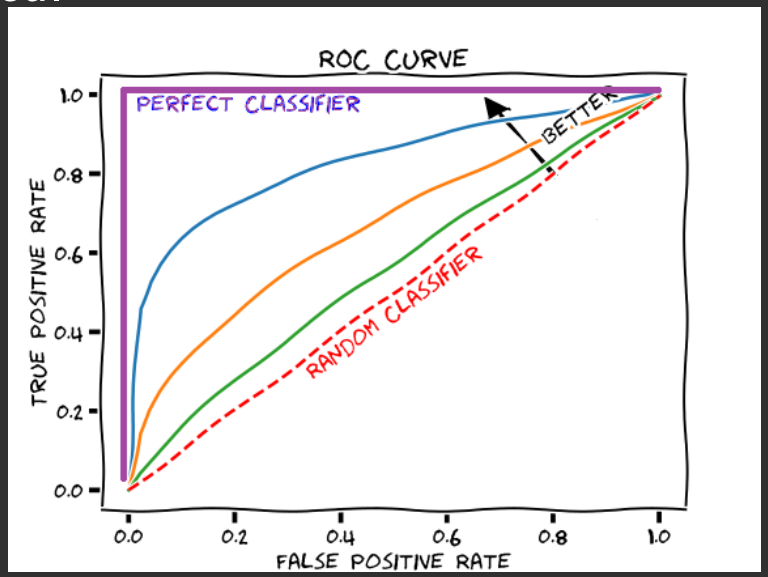
Data generation



E.g. $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\} \Longrightarrow$ Known:$\{0, 1, 2, 3, 4, 5\}$Unknown:$\{6, 7, 8, 9\}$

Evaluation

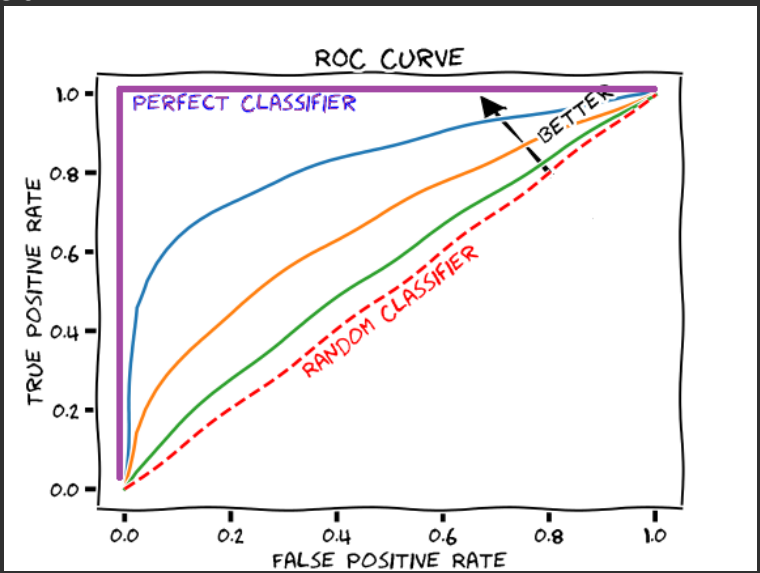
AUROC: A receiver operating characteristic curve, or ROC curve, is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied.



OSCR: The testing accuracy on a mixed testing set (Known and unknown sets are all included). A sample is considered to be classified correct if:
It is from known set, and we labeled it correctly.

Evaluation

AUROC: A receiver operating characteristic curve, or ROC curve, is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied.



OSCR: The testing accuracy on a mixed testing set (Known and unknown sets are all included). A sample is considered to be classified correct if:
It is from known set, and we labeled it correctly.

Methodologies and results

Logistic Regression

Multi-classification

Train one model for each label and output the result with highest probability

Reject unknown

Before giving final result,calculate its score:

$$Z = w^T x + b$$

and

$$f(x) = \begin{cases} \text{Known} & Z > \theta \\ \text{Unknown} & Z < \theta \end{cases} \tag{1}$$

here θ is a hyper-parameter

Nearest Neighbors

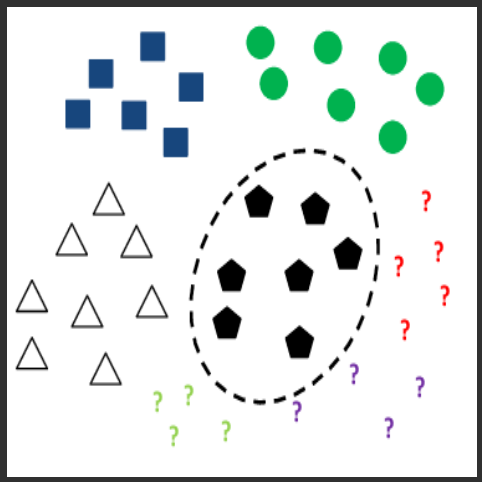


Figure: Decision boundaries of NN method

Classification and reject unknown

Obtains the nearest neighbor t and the second nearest neighbor u with different labels of input data s and get distance ratio:

$$R = d(s,t)/d(s,u)$$

and

$$f(s) = \begin{cases} t\text{'s label} & R < \theta \\ \text{Unknown} & R > \theta \end{cases} \tag{2}$$

here $0 < \theta < 1$ is a hyper-parameter

SVM

Multi-classification

Train a support vector machine with rbf kernels. Out put the probability for each class.

Radial basis function kernel

$$K(x, x') = exp(-\frac{\|x - x'\|^2}{2\sigma^2})$$

Reject unknown

If $maxP_i < P_{thres}$, label it as unknown. Here P_{thres} is a hyper parameter. Otherwise label the sample as $\underset{i}{\operatorname{argmax}} P_i$

CNN

Network structure

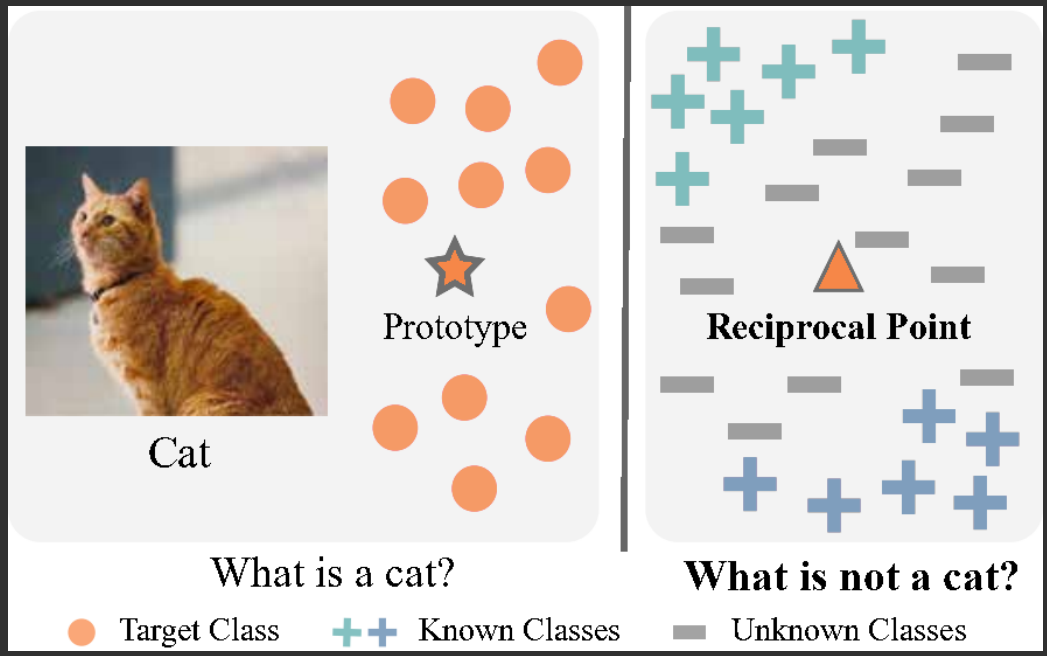
2 convolutional layer,with RELU activation and max pooling. 1 fully connected layer.

Reject unknown

Each class has a softmax probability P_i, If $maxP_i < P_{thres}$, label it as unknown. Here P_{thres} is a hyper parameter. Otherwise label the sample as $\underset{i}{\operatorname{argmax}} P_i$

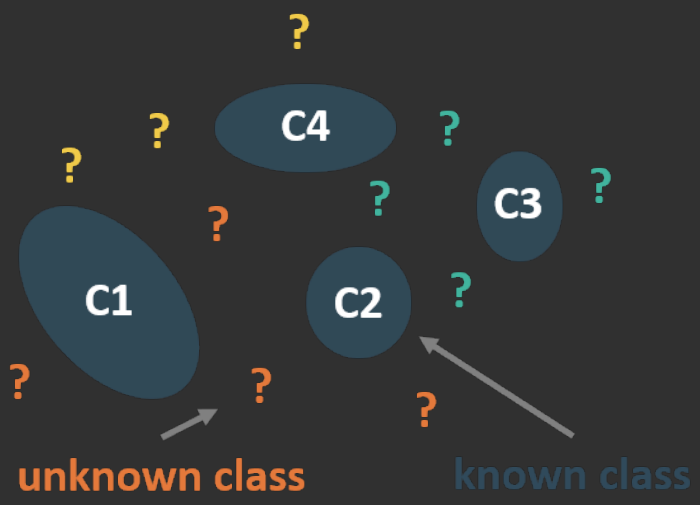
Not Satisfying results. Why?

closed-set: Margin? Prototype?
Only focus on limited parts of the whole open space!
Better way of thinking: How to represent unknown?



Adversarial Reciprocal Points Learning!

Adversarial Reciprocal Points Learning



Reciprocal Points

The *reciprocal point* $\mathcal{P}^k$ of category k is regarded as the latent representation of the all data points that don't belong to category k. Hence, the samples of other parts in open space should be closer to the reciprocal point $\mathcal{P}^k$ than the samples of category k.

Adversarial Reciprocal Points Learning

Reciprocal Points

Specifically, Given an deep embedding function $\mathcal{C}$ with learnable parameters θ, sample x and reciprocal point $\mathcal{P}_k$, their distance $d(\mathcal{C}(x), \mathcal{P}^k)$ is calculated by combining the Euclidean distance d_e and dot product d_d:

$$
\begin{aligned}
d_e(\mathcal{C}(x), \mathcal{P}^k) &= \frac{1}{m} \cdot \|\mathcal{C}(x) - \mathcal{P}^k\|_2^2, \\
d_d(\mathcal{C}(x), \mathcal{P}^k) &= \mathcal{C}(x) \cdot \mathcal{P}^k, \\
d\ (\mathcal{C}(x), \mathcal{P}^k) &= d_e(\mathcal{C}(x), p_i^k) - d_d(\mathcal{C}(x), \mathcal{P}^k).
\end{aligned}
\tag{3}
$$

Adversarial Reciprocal Points Learning

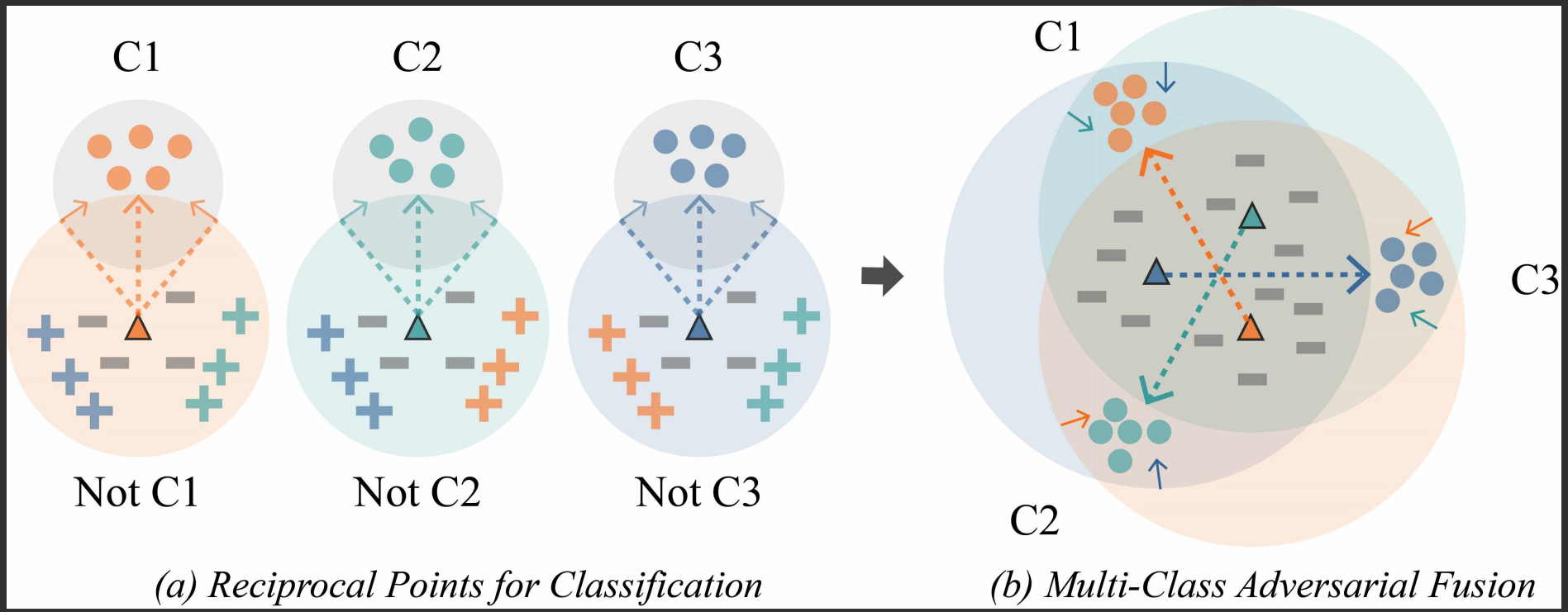
Reciprocal Points

The learning of θ is achieved by minimizing the reciprocal points classification loss based the negative log-probability of the true class k:

$$\mathcal{L}_c(x;\theta,\mathcal{P}) = -\ log\ p(y=k|x,\mathcal{C},\mathcal{P}). \tag{4}$$

Through minimizing Eq. (4), the reciprocal points classification loss reduces the empirical classification risk through the reciprocal points.

Adversarial Reciprocal Points Learning



(a) Reciprocal Points for Classification *(b) Multi-Class Adversarial Fusion*

Each known class should be pushed to the edge of the finite feature space to the maximum extent, moving each far away from its potential unknown space. Adversarial margin is used to solve this problem.

Adversarial Reciprocal Points Learning

Adversarial Margin Constraint

The open space risk can be bounded indirectly by constraining the distance between the samples from class k and the reciprocal points $\mathcal{P}^k$ to be smaller than R as follows:

$$\mathcal{L}_o(x;\theta,\mathcal{P}^k,R^k) = max(d_e(\mathcal{C}(x),\mathcal{P}^k) - R, 0), \tag{5}$$

where R is a learnable margin.

Adversarial Reciprocal Points Learning

Learning framework

The overall loss function combines Eq. (4) and Eq. (5) to handle the empirical classification risk and the open space risk simultaneously:

$$\mathcal{L}(x, y; \theta, \mathcal{P}, R) = \mathcal{L}_c(x; \theta, \mathcal{P}) + \mathcal{L}_o(x; \theta, \mathcal{P}, R), \tag{6}$$

where $\theta, \mathcal{P}, R$ represent the learnable parameters.

Results

Table 1: The AUROC results of on detecting known and unknown samples. Results are averaged among five randomized trials.

Method	MNIST	CIFAR10	CIFAR100-50
Logistic	83.57 ±0.1	...	...
SVM	93.50 ±0.1	...	...
NN	89.12 ±0.1	...	...
CNN	90.70 ±0.2	...	...
ARPL	99.34 ± 0.1	89.01 ± 0.3	91.20 ± 0.2

Results

Table 2: The open set classification rate (OSCR) curve results of open set recognition. Results are averaged among five randomized trials.

Method	MNIST	CIFAR10	CIFAR100-50
Logistic	79.85 ±0.1	...	...
SVM	90.05 ±1.0	...	...
NN	89.58 ±0.1	...	...
CNN	92.00 ±0.5	...	...
ARPL	99.18 ±0.1	85.20 ±0.7	89.74 ±0.5

Conclusion

This project

- Modify several algorithms (Logistic, Nearsest Neighbours, SVM and CNN) to solve *Open-Set-Recognition* problems.
- Compare and analyze the results, to achieve better accuracy, the overall open space needs to be considered when training.
- Reproduce the method ARPL to solve above queations by learning latent representation of unknown and set retrictions to seperate them with known samples.

Future work

Datsets

- By now most methods are tested with relative small datasets, like MNIST, CIFAR and TinyImageNet. Much bigger datasets need to be taken into experiments.
- Most existing methods do not have a clear definition of 'semantic class' which OSR should be focus on. A semantic shift (more fine-grained)benchmark was proposed to better understand and evaluate OSR.[1]

[1] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: a good closed-set classifier is all you need In *CVPR*, 2021.

Fine-grained Dataset

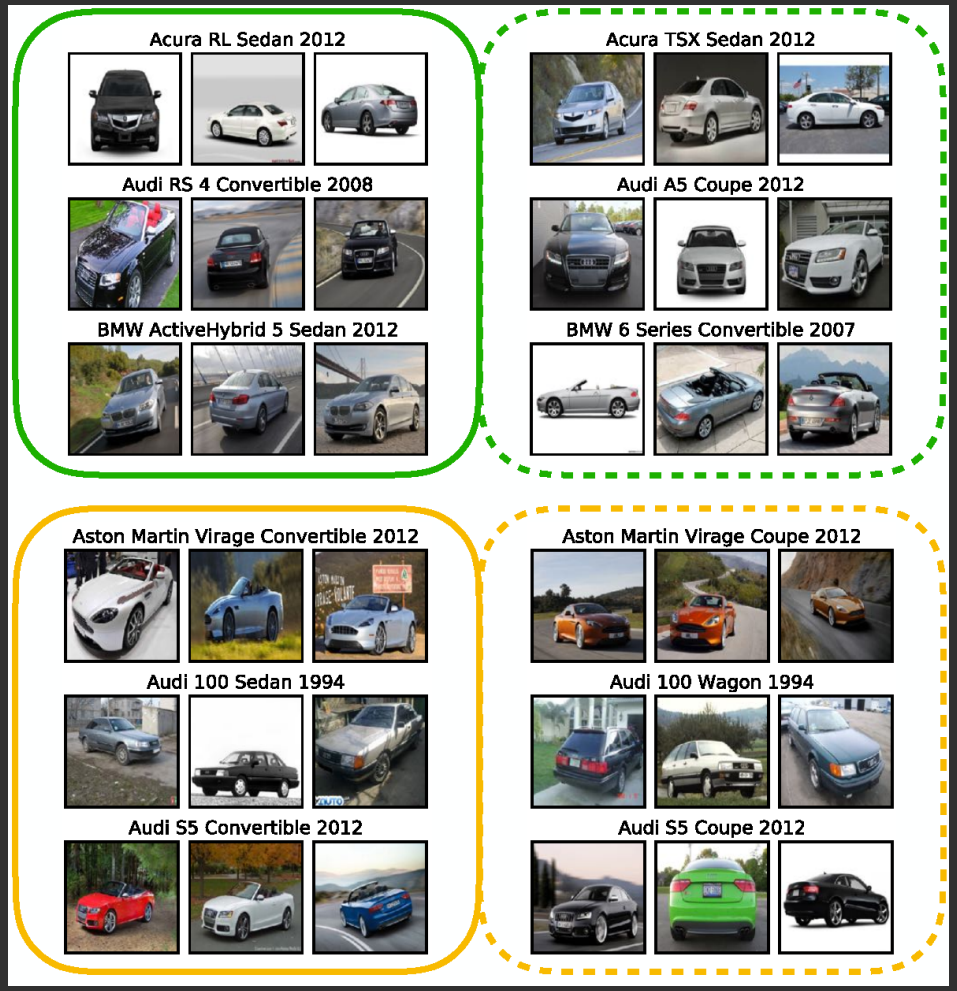


Figure: Scars Example[2]

[2] https://www.kaggle.com/datasets/jessicali9530/stanford-cars-dataset

The End