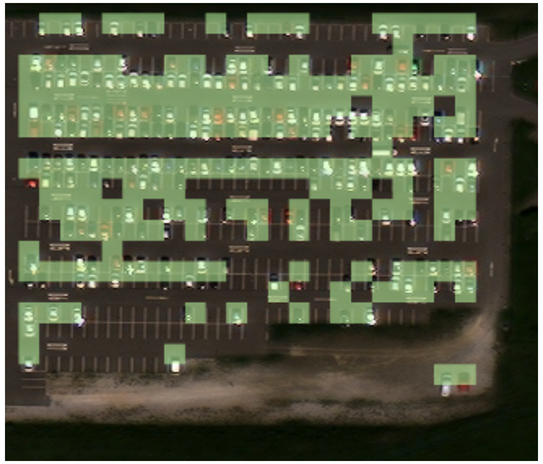
Guided Generative Models using Weak Supervision for Detecting Object Spatial Arrangement in Overhead Images

Presenter: Weiwei Duan

University of Southern California

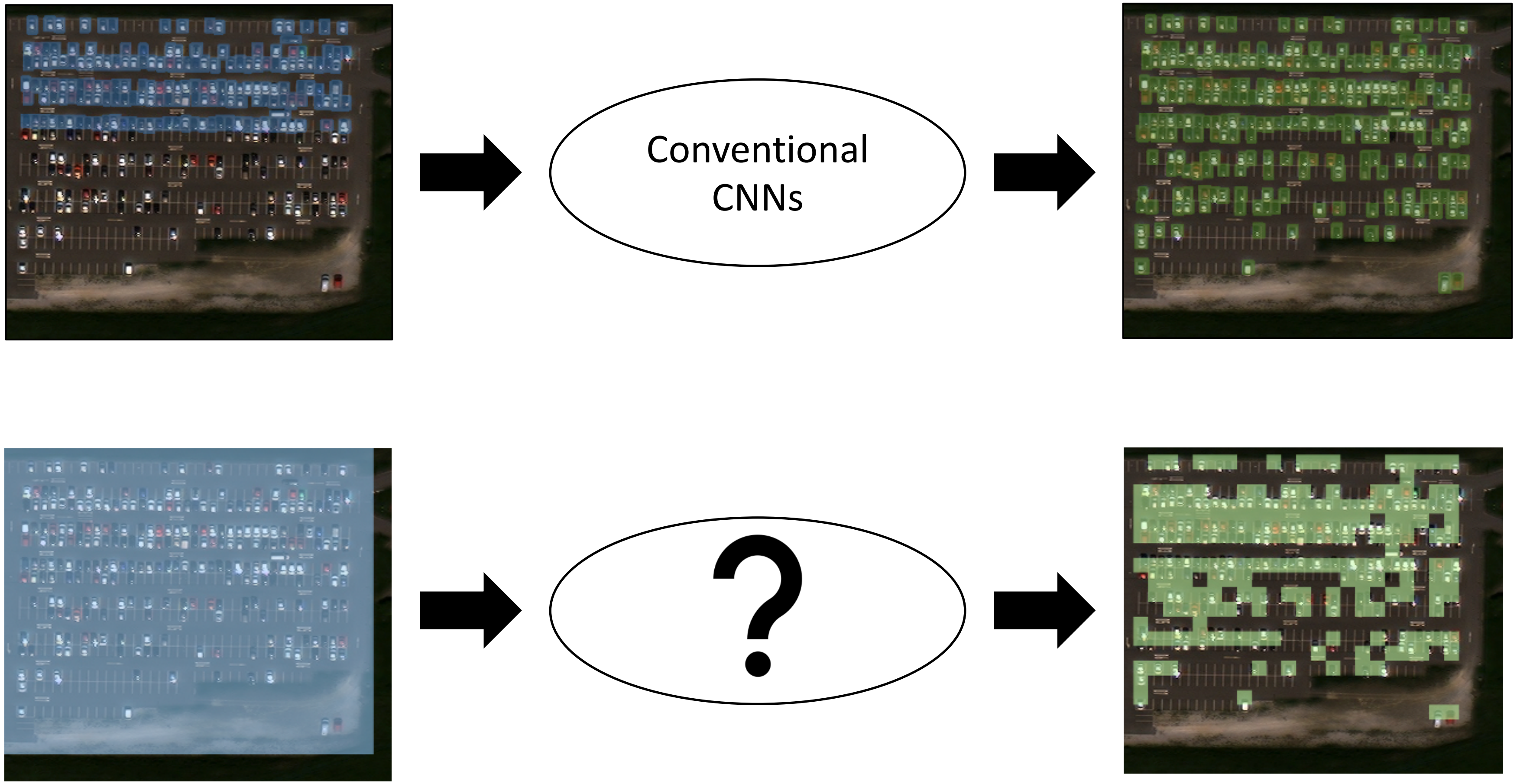
Spatial Arrangement

- The location of a group of desired objects
- E.g., most of cars are parked in the northside of the parking lot



Motivation

- Leverage the Region-of-interest to obtain coarse but useful results

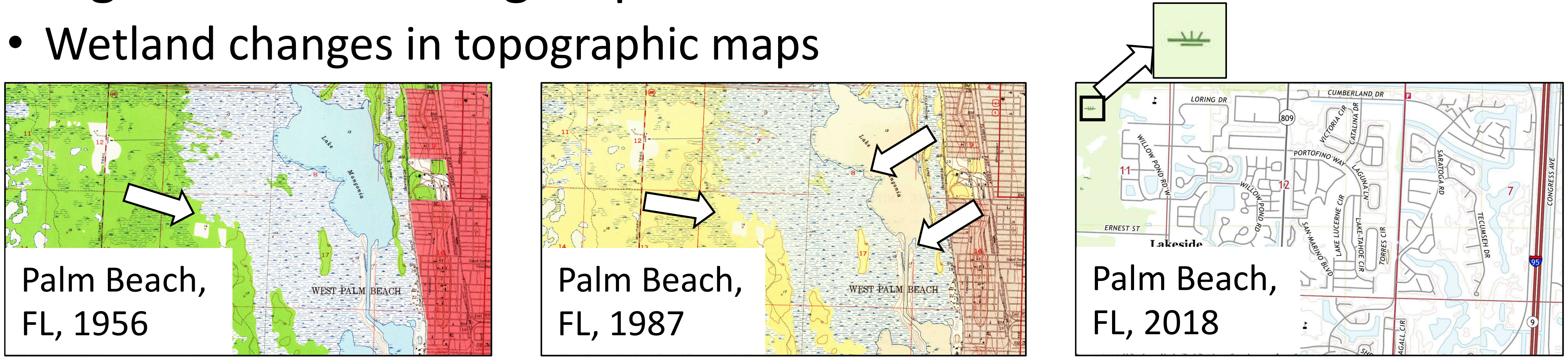


Coarse but Useful

- Crowd surveillance for disease monitoring
  - When was covid-19 active in China [1]

[1] Nsoesie, Elaine Okanyene, et al. "Analysis of hospital traffic and search engine data in Wuhan China indicates early disease activity in the Fall of 2019." (2020).

- Changes detection for geospatial features
  - Wetland changes in topographic maps

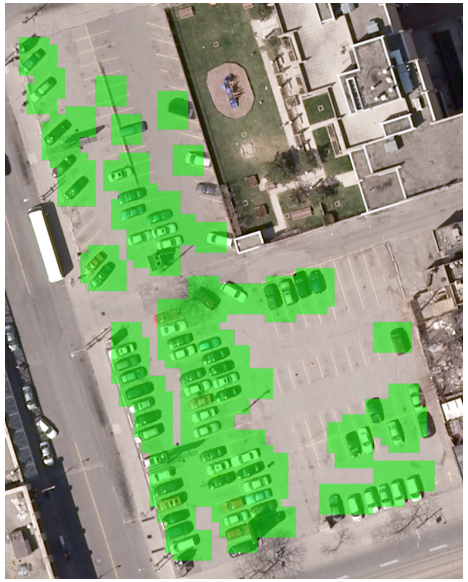


Problem Statement

- Detect the **approximate locations** of a group of target objects in a region-of-interest (**ROI**) in overhead images
- **Manual work:** label one or a few target sample(s)

The weak annotation
(obtained from external datasets
Or manual labeling)

Inputs



Outputs

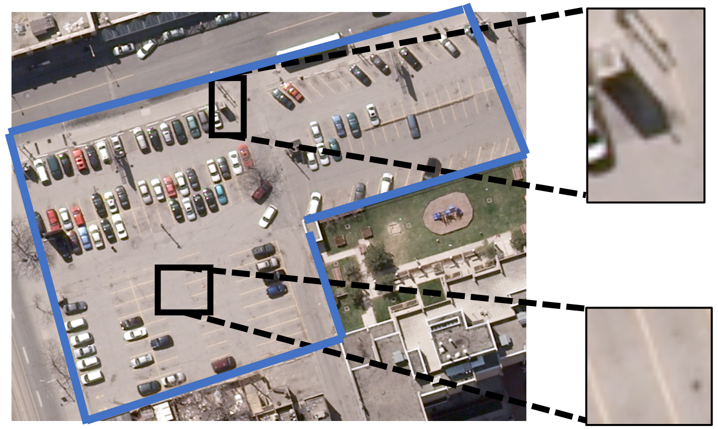
Challenges

- No sufficient labeled samples to cover the diversity of targets
- No labeled samples for non-target objects

V.S.

Labeled target

Unlabeled targets

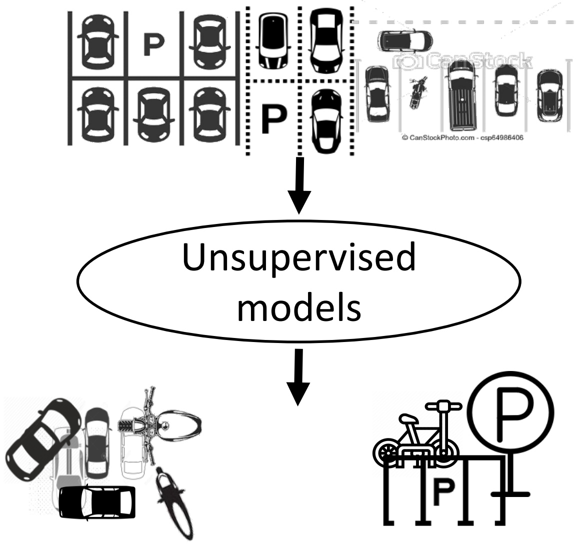


Non-targets in the region

Existing Work

	Labeled targets	Labeled non-targets	Results accuracy
Unsupervised [2,3]			Low
Semi-supervised [4]			
Our model			

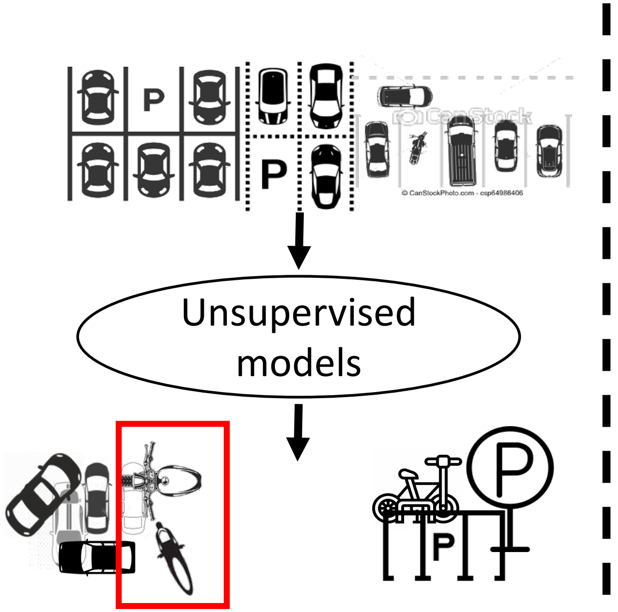
[2] Yang et al., 2019
[3] Jiang et al., 2016
[4] Zhang et al., 2019



Existing Work

	Labeled targets	Labeled non-targets	Results accuracy
Unsupervised [2,3]			Low
Semi-supervised [4]			
Our model			

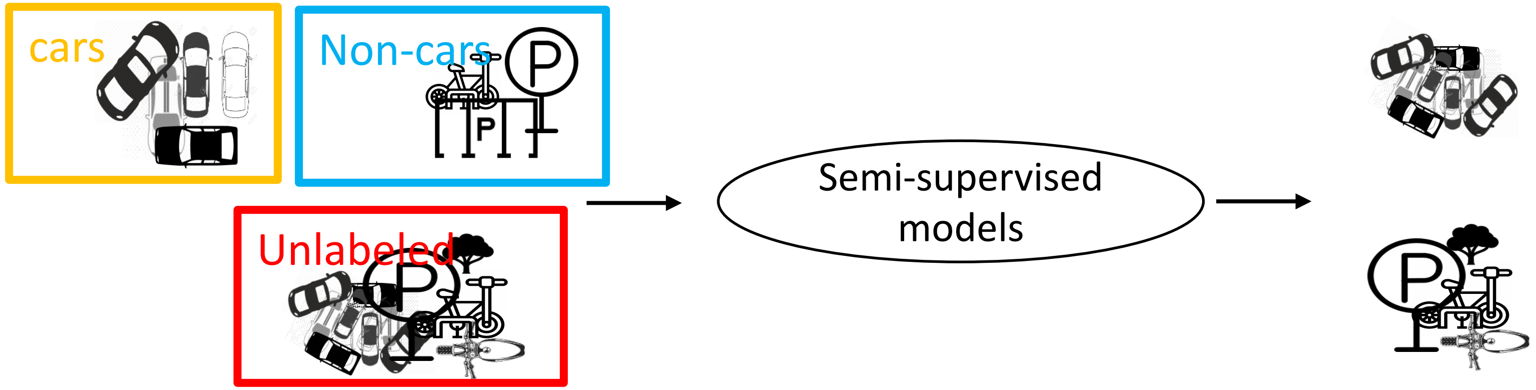
[2] Yang et al., 2019
[3] Jiang et al., 2016
[4] Zhang et al., 2019



Existing Work

	Labeled targets	Labeled non-targets	Results accuracy
Unsupervised [2,3]			Low
Semi-supervised [4]			High
Our model			High

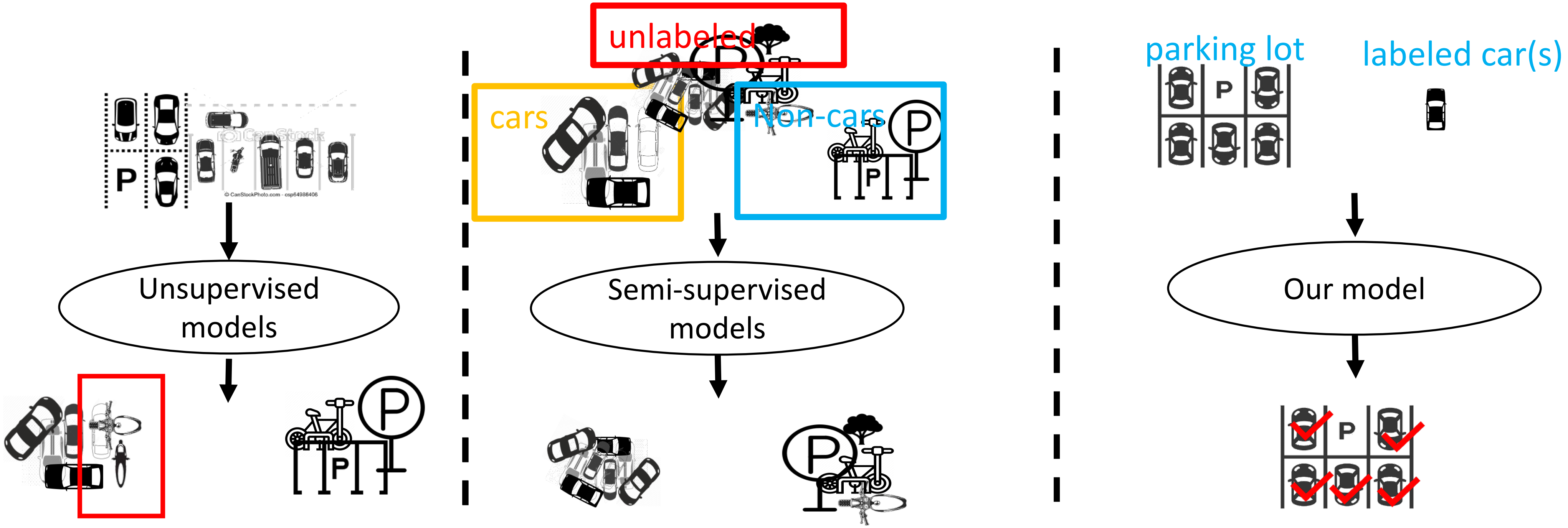
[2] Yang et al., 2019
[3] Jiang et al., 2016
[4] Zhang et al., 2019



Existing Work

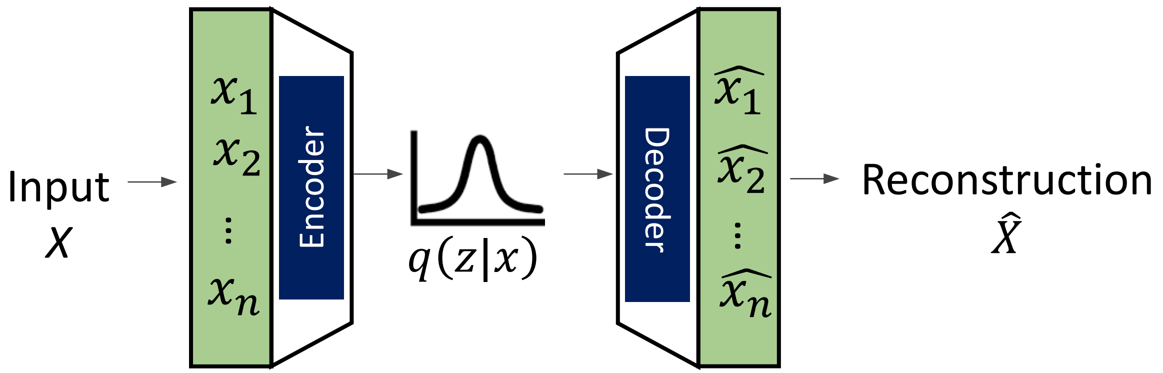
	Labeled targets	Labeled non-targets	Results accuracy
Unsupervised [1,2]			Low
Semi-supervised [3]			High
Our model			High

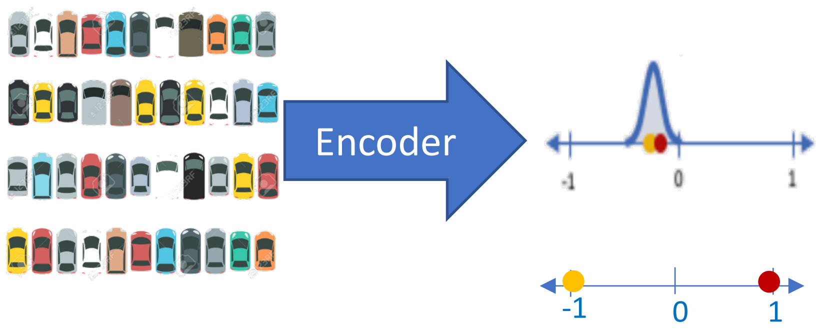
[1] Jiang et al., 2016
[2] Dilokthanakul et al., 2016
[3] Maaløe et al., 2019



Generative Models for Scarce Labels

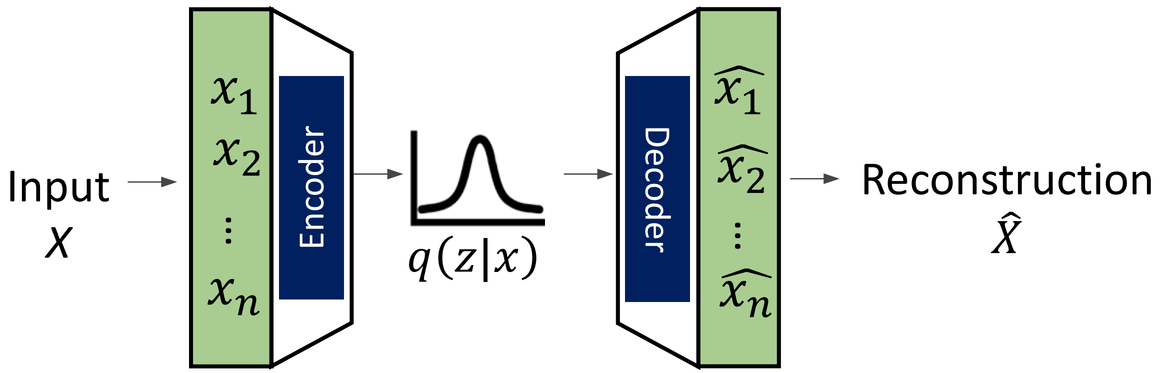
- Variational Auto-encoder (VAE), generative models
  - Learning representation distribution, *z*

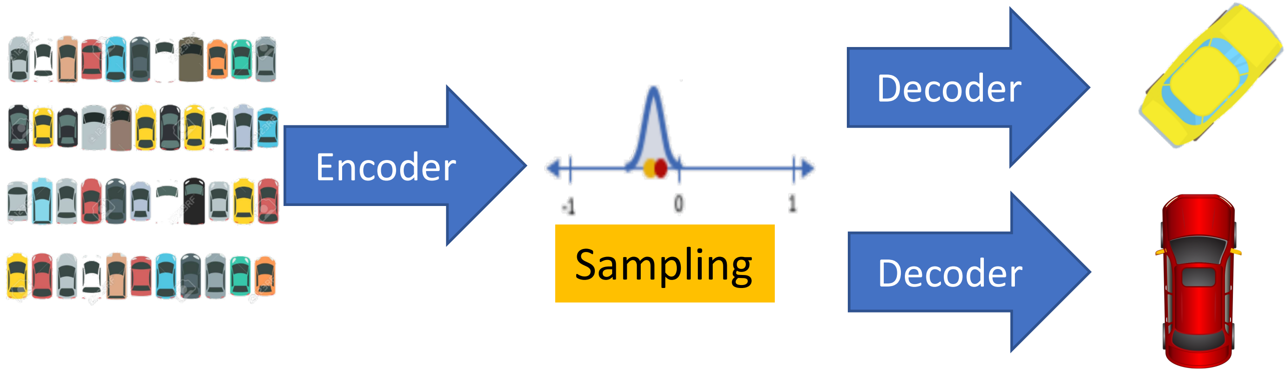




Generative Models for Scarce Labels

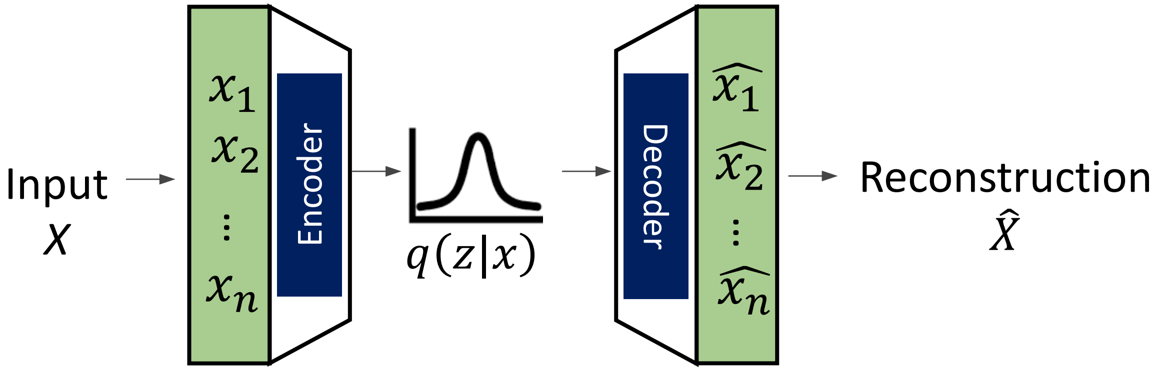
- Variational Auto-encoder (VAE)
  - Learning representation distribution, *z*

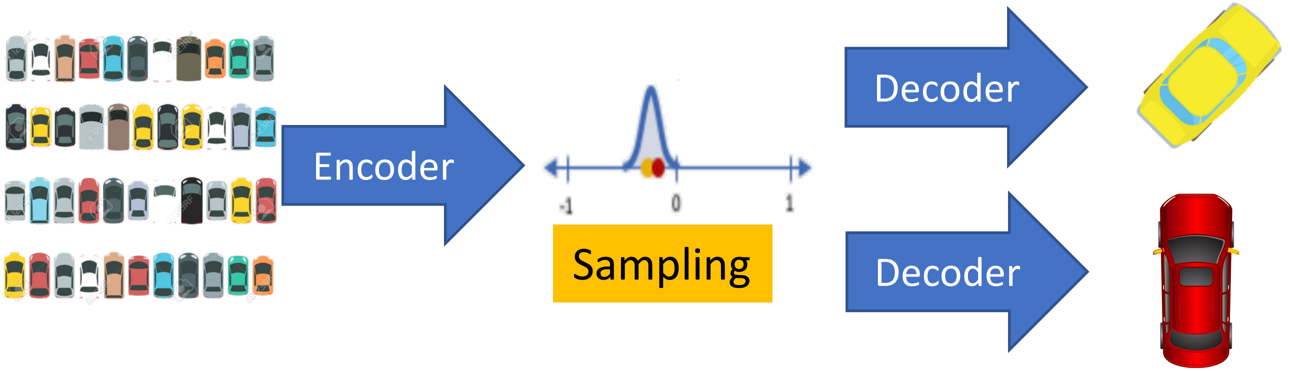




Generative Models for Scarce Labels

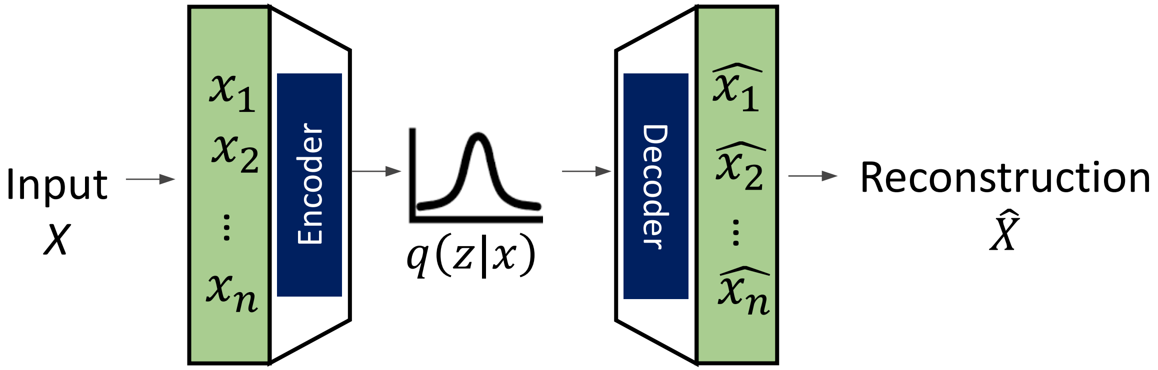
- Variational Auto-encoder (VAE)
  - Learning representation distribution, *z*

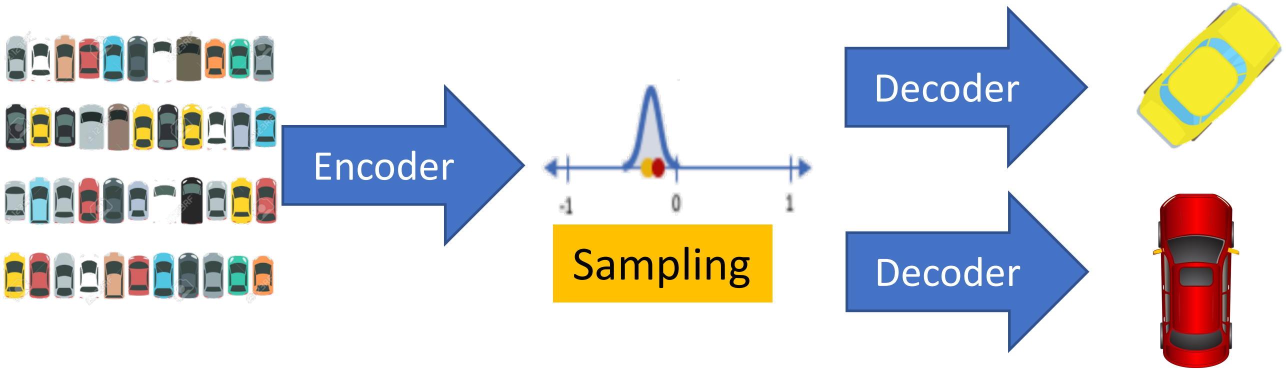




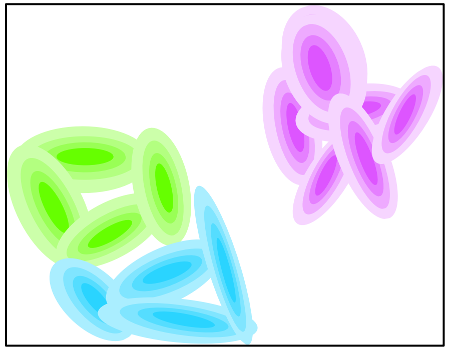
Generative Models for Scarce Labels

- Variational Auto-encoder (VAE)
  - Learning representation distribution, *z*



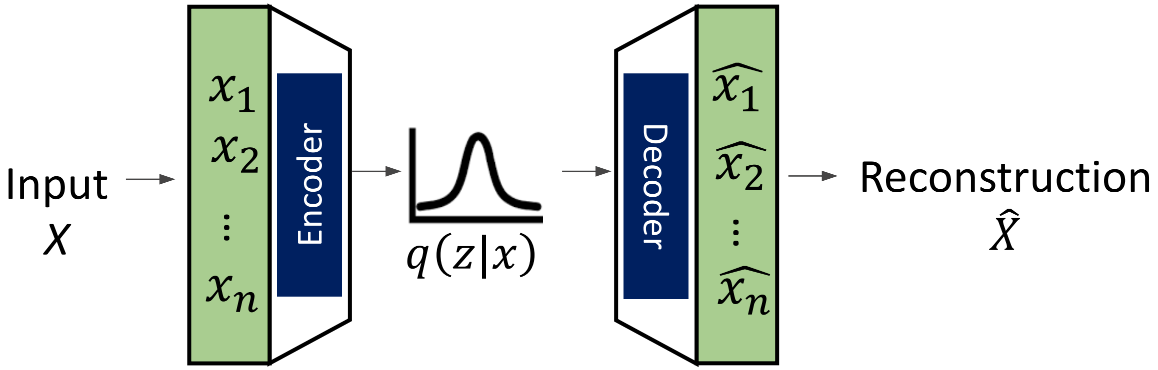


Feature representations from VAE

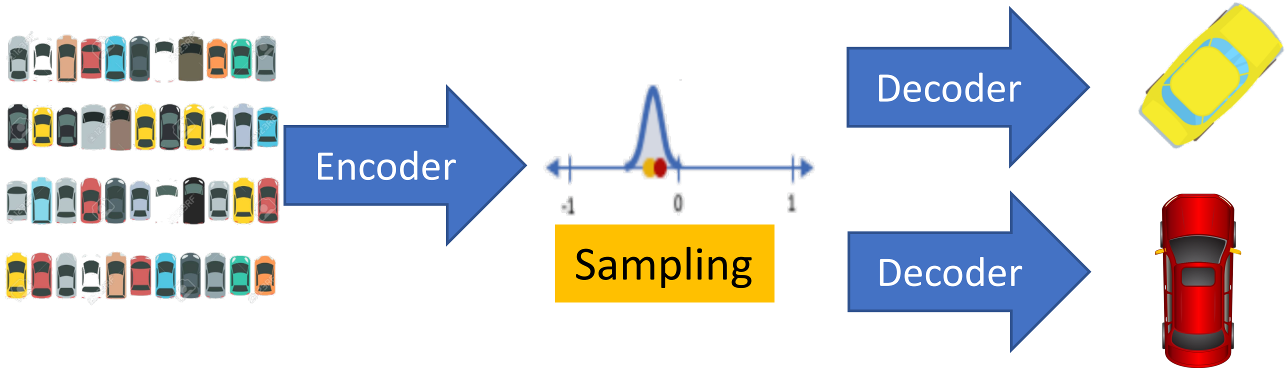


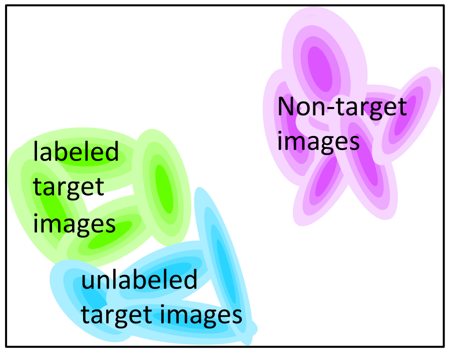
Generative Models for Scarce Labels

- Variational Auto-encoder (VAE)
  - Learning representation distribution, *z*



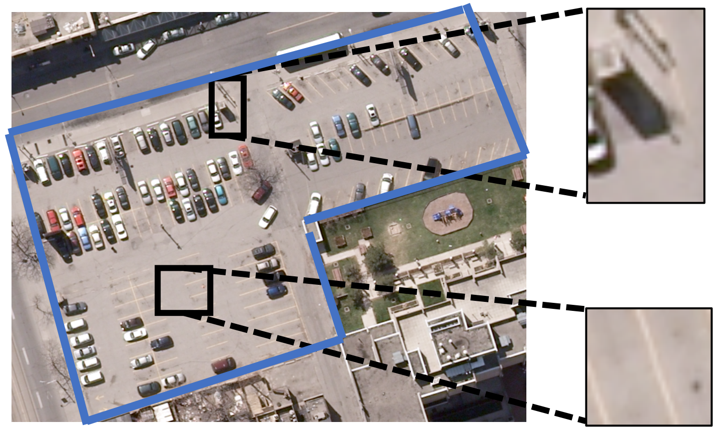
Feature representations from VAE



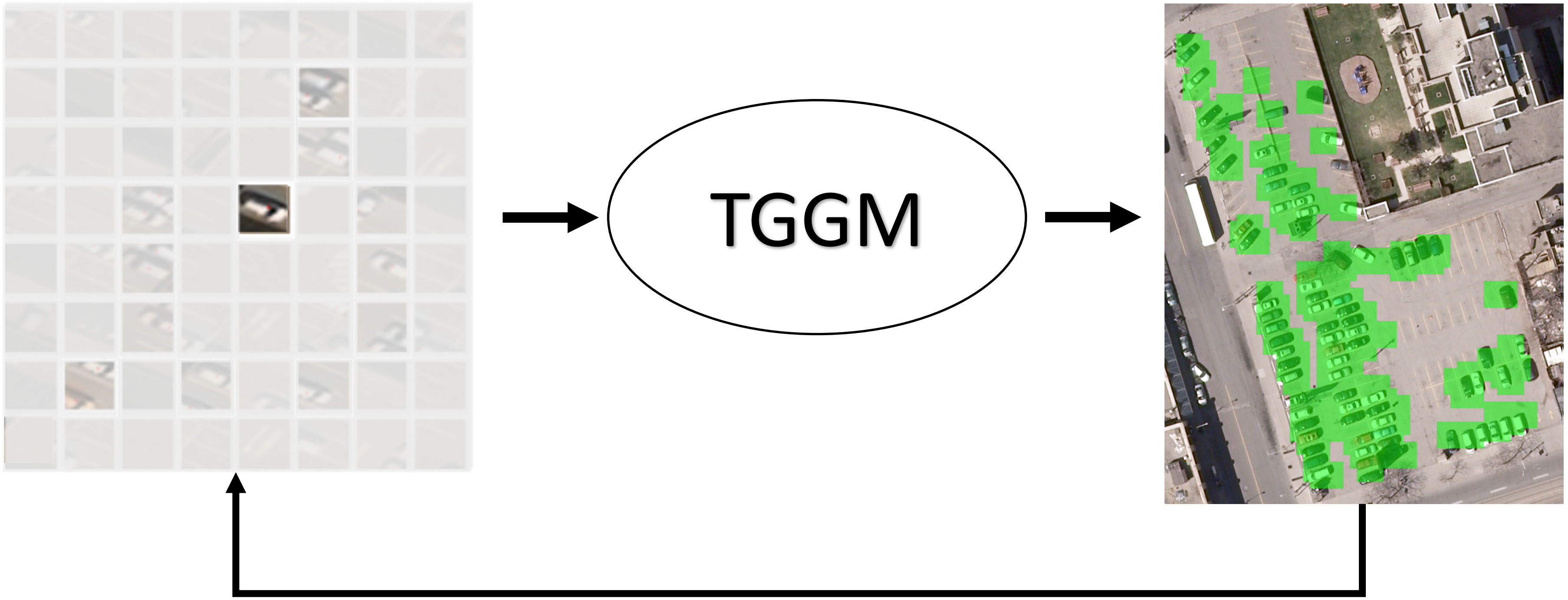


Solution to No Labeled Non-target Images

- The ROI have strong semantic relationship with the target objects
  - The parking lots and cars
- The non-target objects in the region-level annotations are similar
- Limited variations of non-target objects
- Efficiently classify images into the target and non-target categories



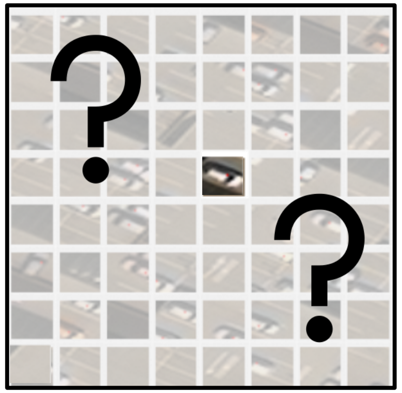
Target-guided Generative Clustering Model



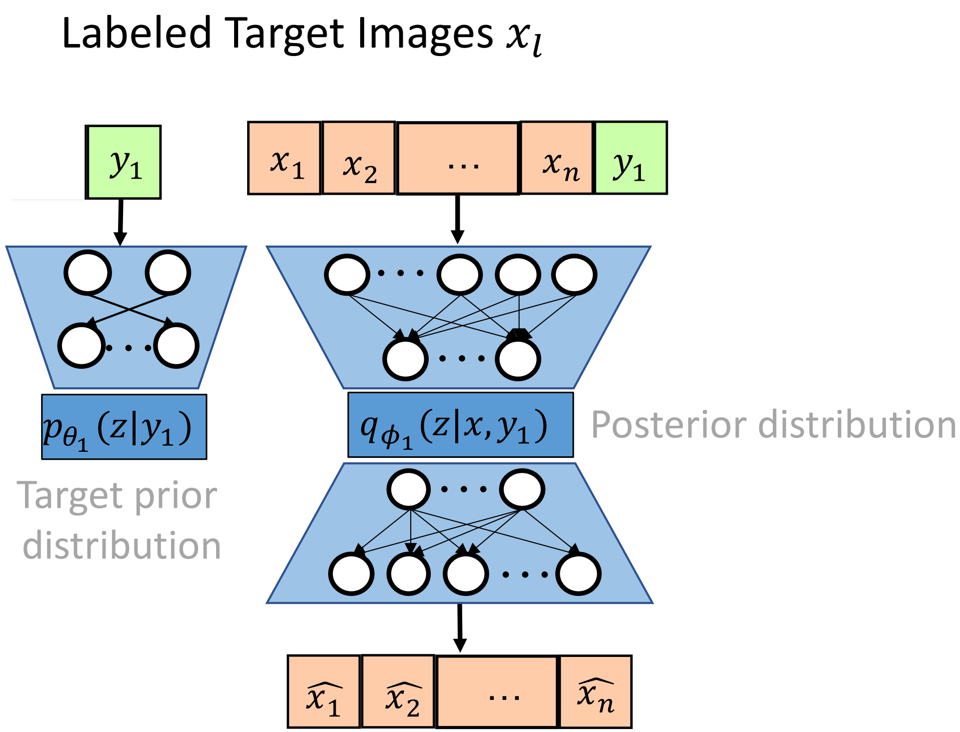
Clear images: labeled or classified target images
Blurred images: unlabeled images

Notation Definitions for TGGM

Notation	Description
x_l	A labeled image covering the target object(s)
x_u	An unlabeled image (target or non-target image)
z	A continuous variable in the hidden space
y	a categorical variable representing an image's label $y = \{y_1, y_2\}$ y_1 for target category and y_2 for non-target category



Target-Guided Generative Clustering Model

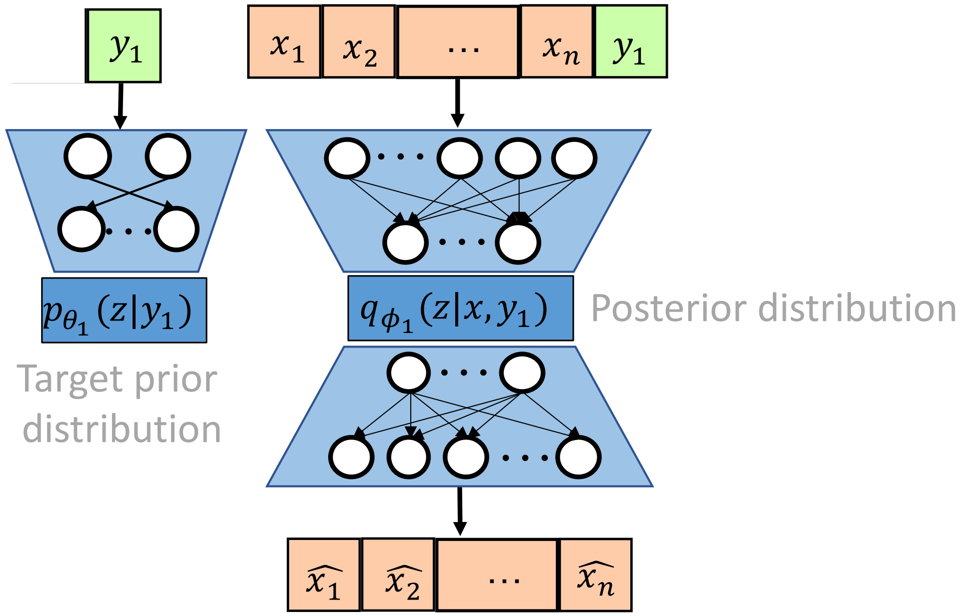


- Two learning goals
  - Output ($\hat{x}$) is similar to the input image (x)
  - The feature distribution is for target images

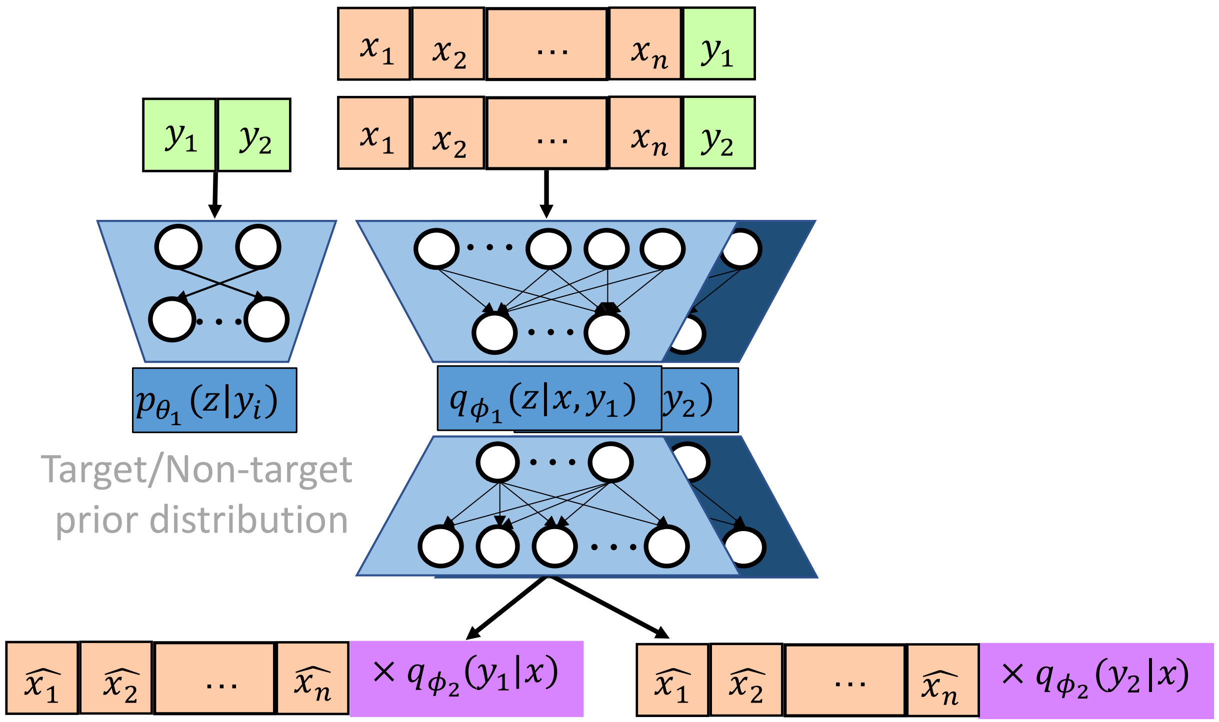
$$Loss(x_t) = \underbrace{MSE(x_l, \widehat{x_l})}_{\text{Reconstruction error}} - \underbrace{KL(q(z|x_l, y_1)||p(z|y_1))}_{\text{Distribution loss}}$$

Target-Guided Generative Clustering Model

Labeled Target Images x_l



Unlabeled Images x_u (including target and non-target images)



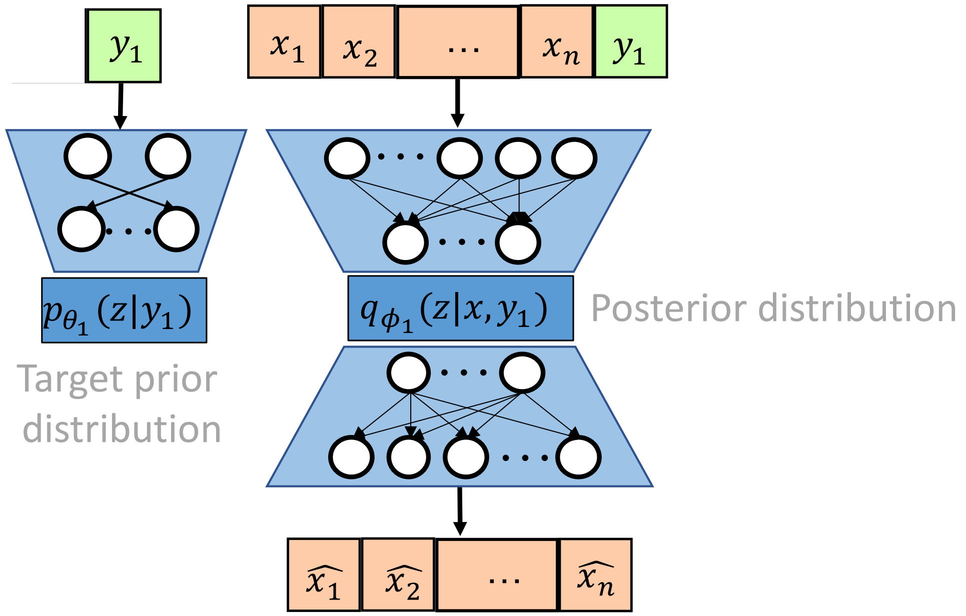
$$Loss(x_t) = MSE(x_l, \widehat{x_l}) - KL(q(z|x_l, y_1)||p(z|y_1))$$

Reconstruction error

Distribution distance

Target-Guided Generative Clustering Model

Labeled Target Images x_l

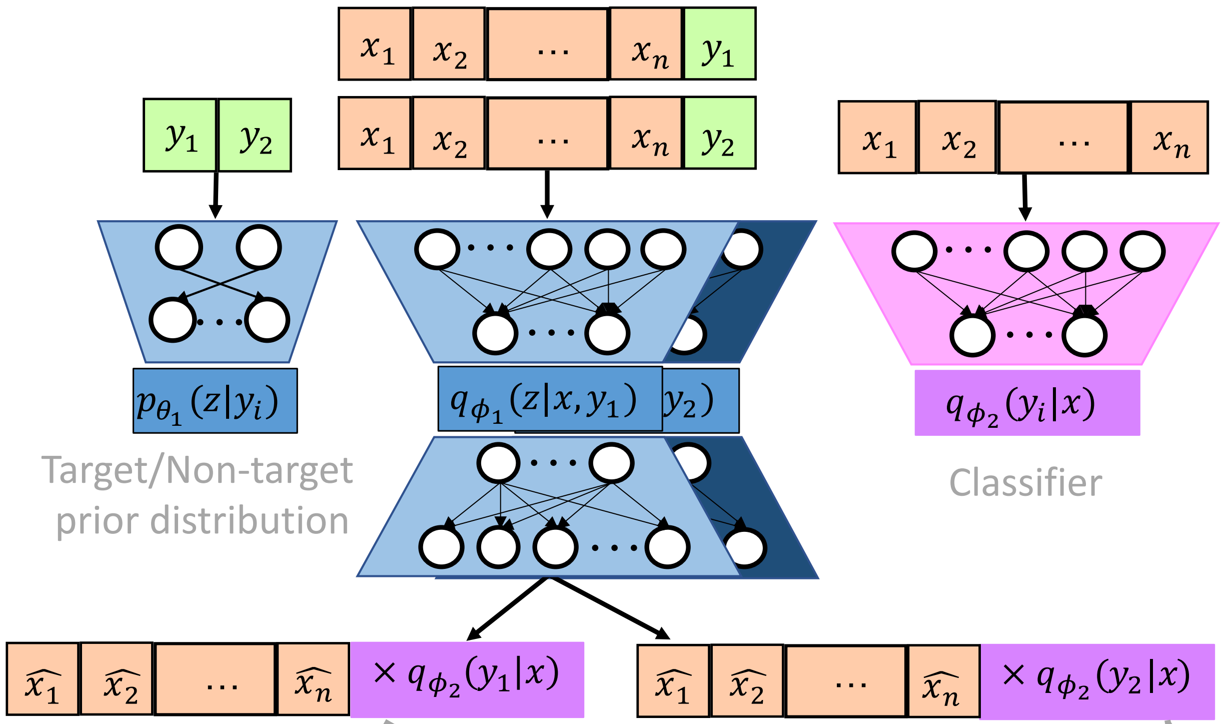


$$Loss(x_t) = MSE(x_l, \widehat{x_l}) - KL(q(z|x_l, y_1)||p(z|y_1))$$

Reconstruction error

Distribution error

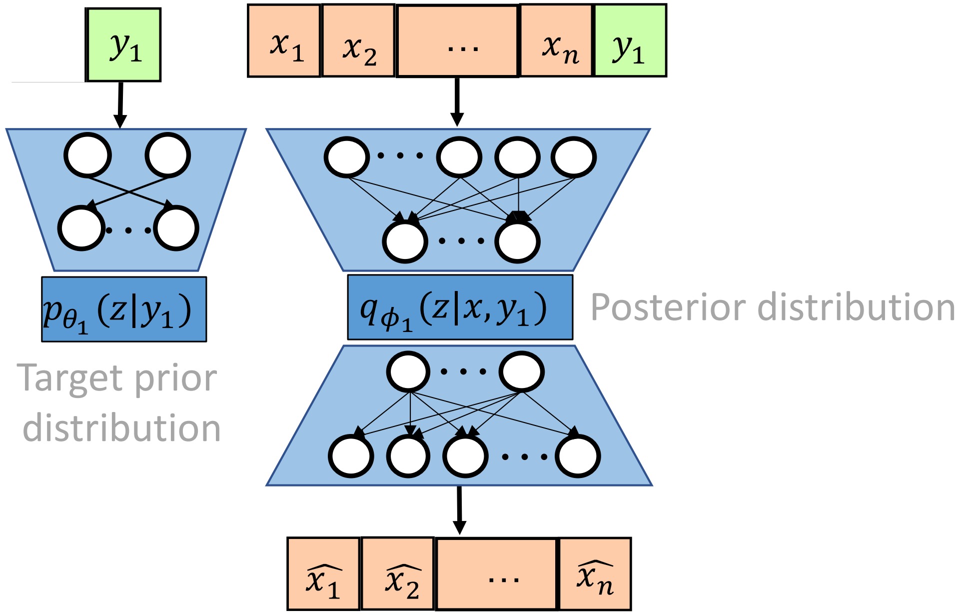
Unlabeled Images x_u (including target and non-target images)



$$Loss(x_u) = q(y_1|x_u) * [MSE(x_u, \widehat{x_{u1}}) - KL(q(z|x_u, y_1)||p(z|y_1))]$$
$$+q(y_2|x_u) * [MSE(x_u, \widehat{x_{u2}}) - KL(q(z|x_u, y_2)||p(z|y_2))]$$
$$-KL[(q(y|x_u)||p(y))]$$

Target-Guided Generative Clustering Model

Labeled Target Images x_l

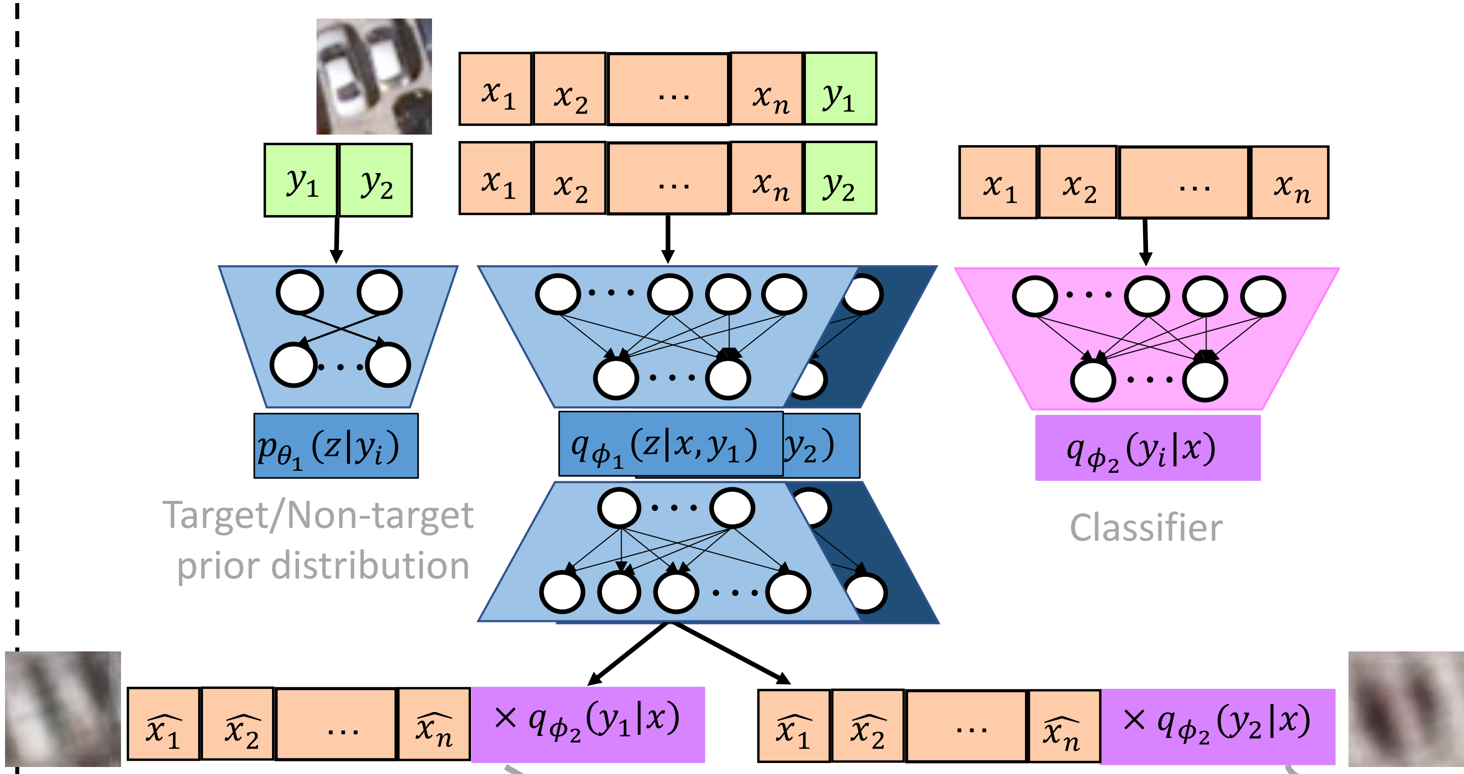


$$Loss(x_t) = MSE(x_l, \widehat{x_l}) - KL(q(z|x_l, y_1)||p(z|y_1))$$

Reconstruction error

Distribution error

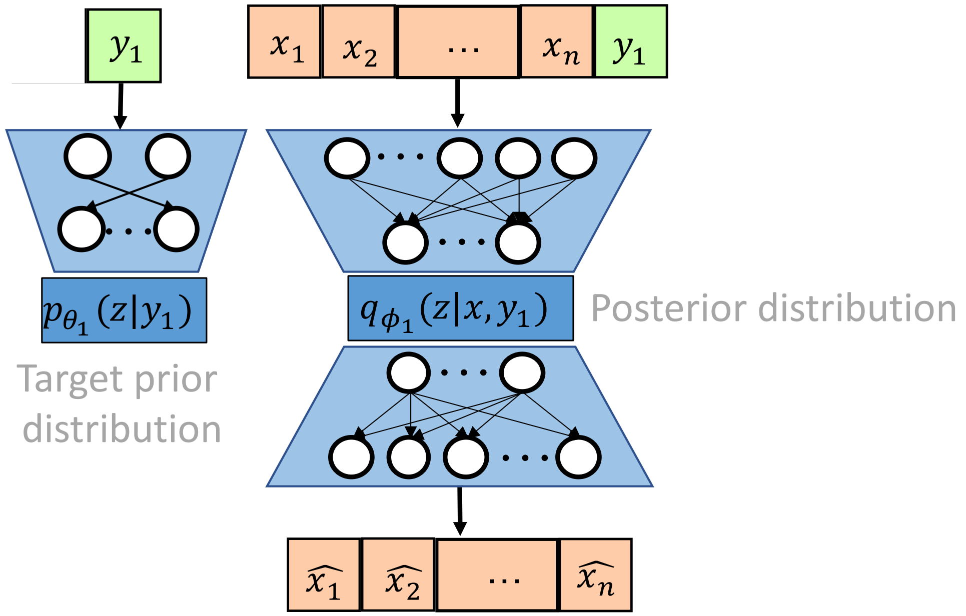
Unlabeled Images x_u (including target and non-target images)



$$Loss(x_u) = q(y_1|x_u) * [MSE(x_u, \widehat{x_{u1}}) - KL(q(z|x_u, y_1)||p(z|y_1))]$$
$$+ q(y_2|x_u) * [MSE(x_u, \widehat{x_{u2}}) - KL(q(z|x_u, y_2)||p(z|y_2))]$$
$$- KL[(q(y|x_u)||p(y))]$$

Target-Guided Generative Clustering Model

Labeled Target Images x_l

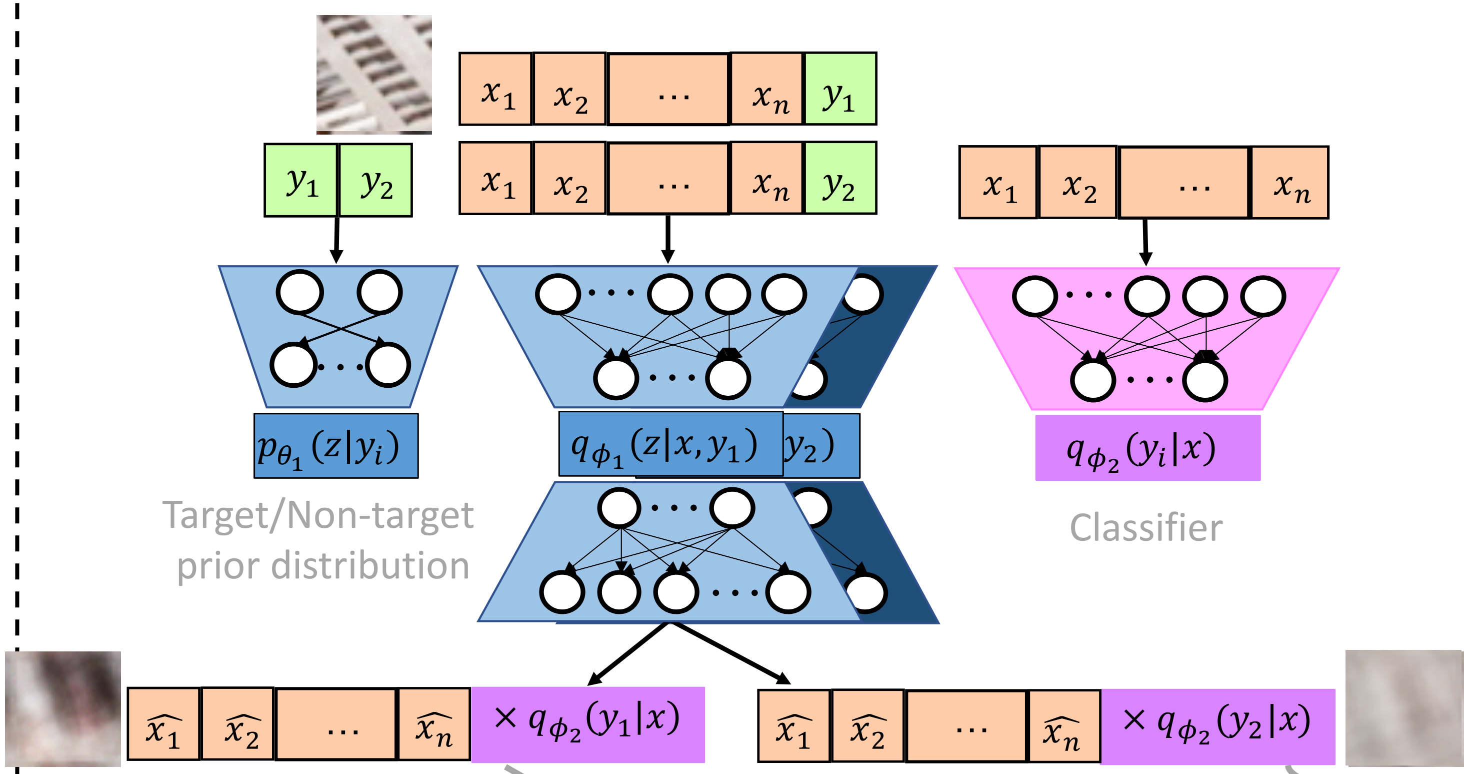


$$Loss(x_t) = MSE(x_l, \widehat{x_l}) - KL(q(z|x_l, y_1)||p(z|y_1))$$

Reconstruction error

Distribution distance

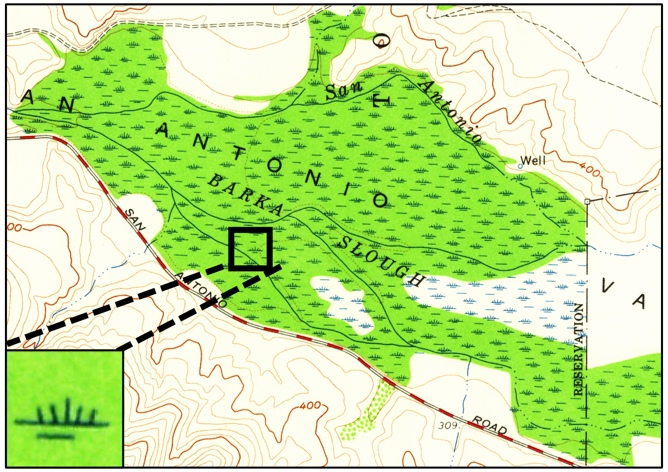
Unlabeled Images x_u (including target and non-target images)



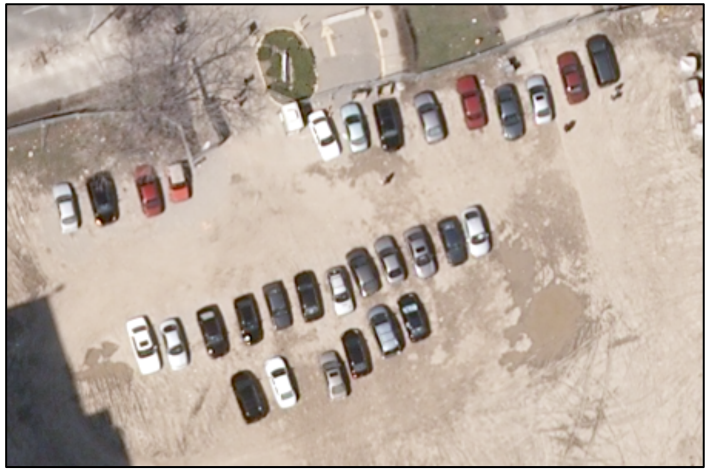
$$Loss(x_u) = q(y_1|x_u) * [MSE(x_u, \widehat{x_{u1}}) - KL(q(z|x_u, y_1)||p(z|y_1))]$$
$$q(y_2|x_u) * [MSE(x_u, \widehat{x_{u2}}) - KL(q(z|x_u, y_2)||p(z|y_2))]$$
$$-KL[(q(y|x_u)||p(y))]$$

Experiment Data

- Satellite imagery
  - Cars Overhead With Context (COWC)
  - Cars and airplanes in xView
  - Ships and airplanes in DIOR
- Topographic maps
  - Wetland areas in USGS topographic maps



USGS



COWC

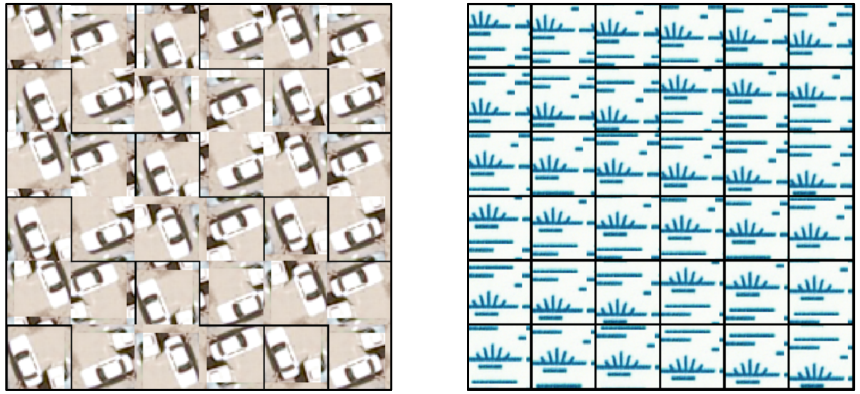
xView

Experiment Settings

Model	Category	Labeled targets	Labeled non-targets	Datasets
TGGM	Weak-supervised	Aug		C, D, U, X
dualAE [2]	Unsupervised			C, D, U, X
VaDE [3]	Unsupervised			C, D, U, X
AVAE [4]	Semi-supervised	40%	40%	C, D, U, X
Yolov3 [5]	supervised	50%	50%	D

[2] Yang et al., 2019
[3] Jiang et al., 2016
[4] Zhang et al., 2019
[5] Redmon et al., 2018

- Aug : augmented labeled target windows
- C, D, U, X : COWC, DIOR, USGS, xView datasets



Augmented target windows

Evaluation Metrics

Model	Category	Labeled targets	Labeled non-targets	Datasets
TGGM	Weak-supervised	Aug		C, D, U, X
dualAE	Unsupervised			C, D, U, X
VaDE	Unsupervised			C, D, U, X
AVAE	Semi-supervised	40%	40%	C, D, U, X
Yolov3	supervised	50%	50%	D

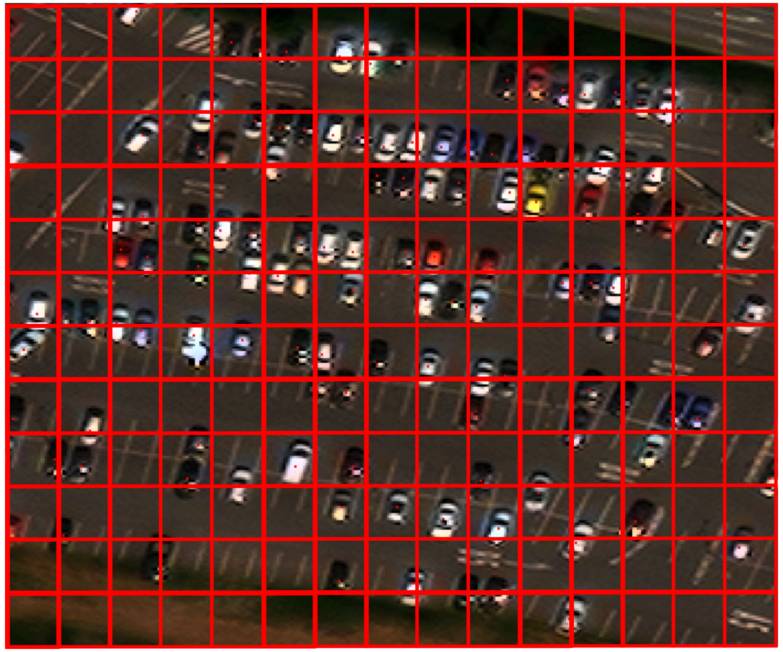
[2] Yang et al., 2019
[3] Jiang et al., 2016
[4] Zhang et al., 2019
[5] Redmon et al., 2018

- First group of experiments
- Evaluate the spatial arrangement estimation
- Precision, recall and F_1 score at the grid-cell level

- Second group of experiments
- Compare with the supervised object detector
- mean Average Precision (mAP)

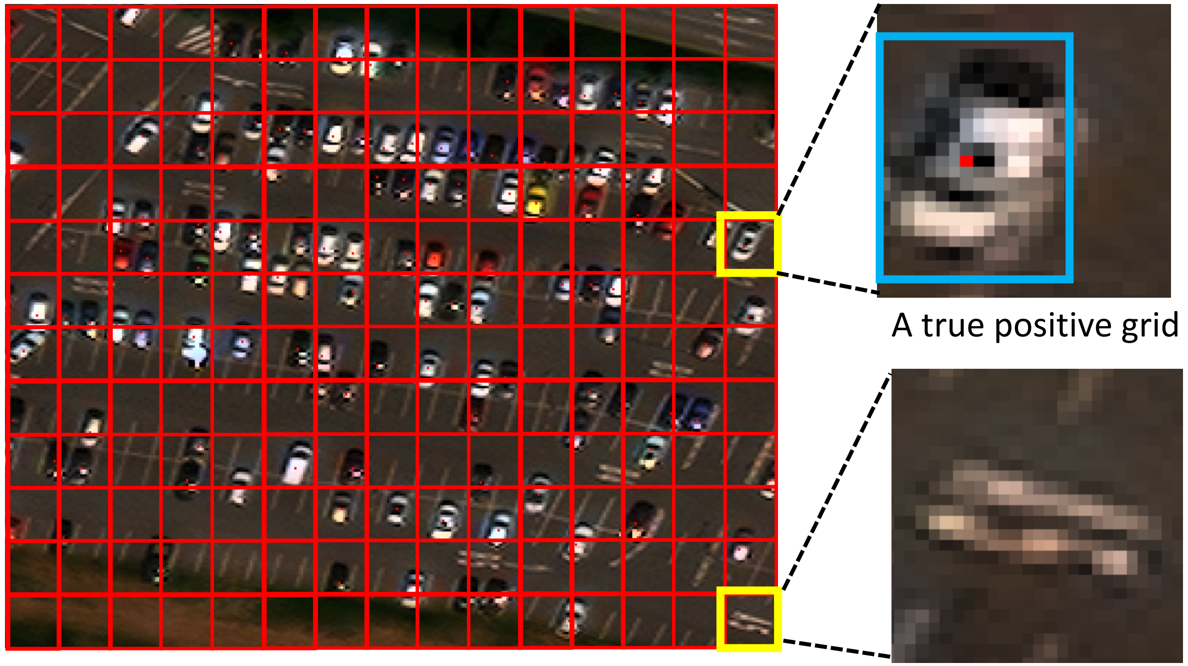
The Grid-cell Level Evaluation

- Slice an image into grid cells (red)



Grid-cell level Evaluation

- Slice an image into grid cells (red)
- True positive grid cells: intersection(grid, bbx) >=0.5
- Otherwise, the grid cells are true negative



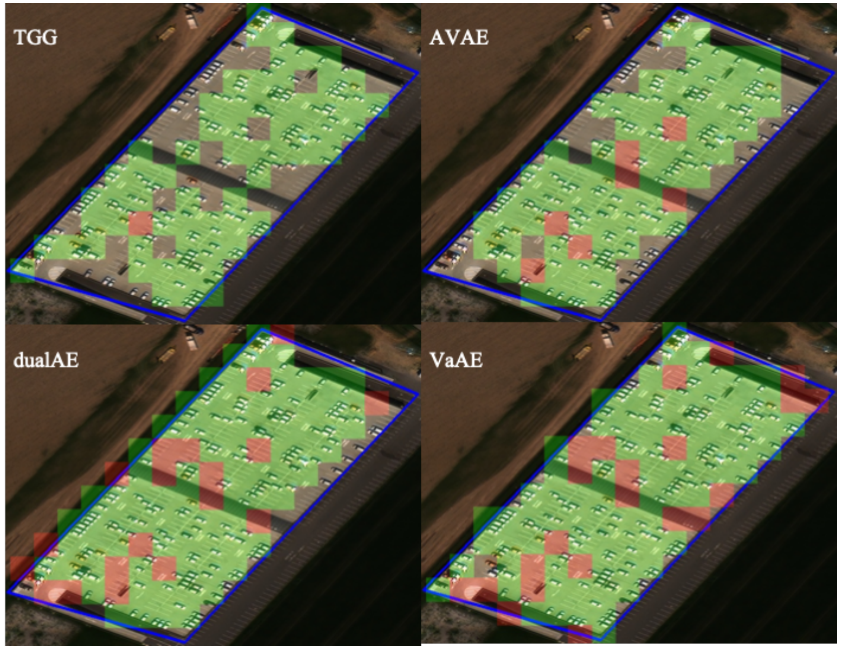
A true positive grid

A true negative grid

Experiment Results

compared with unsupervised and semi-supervised baselines

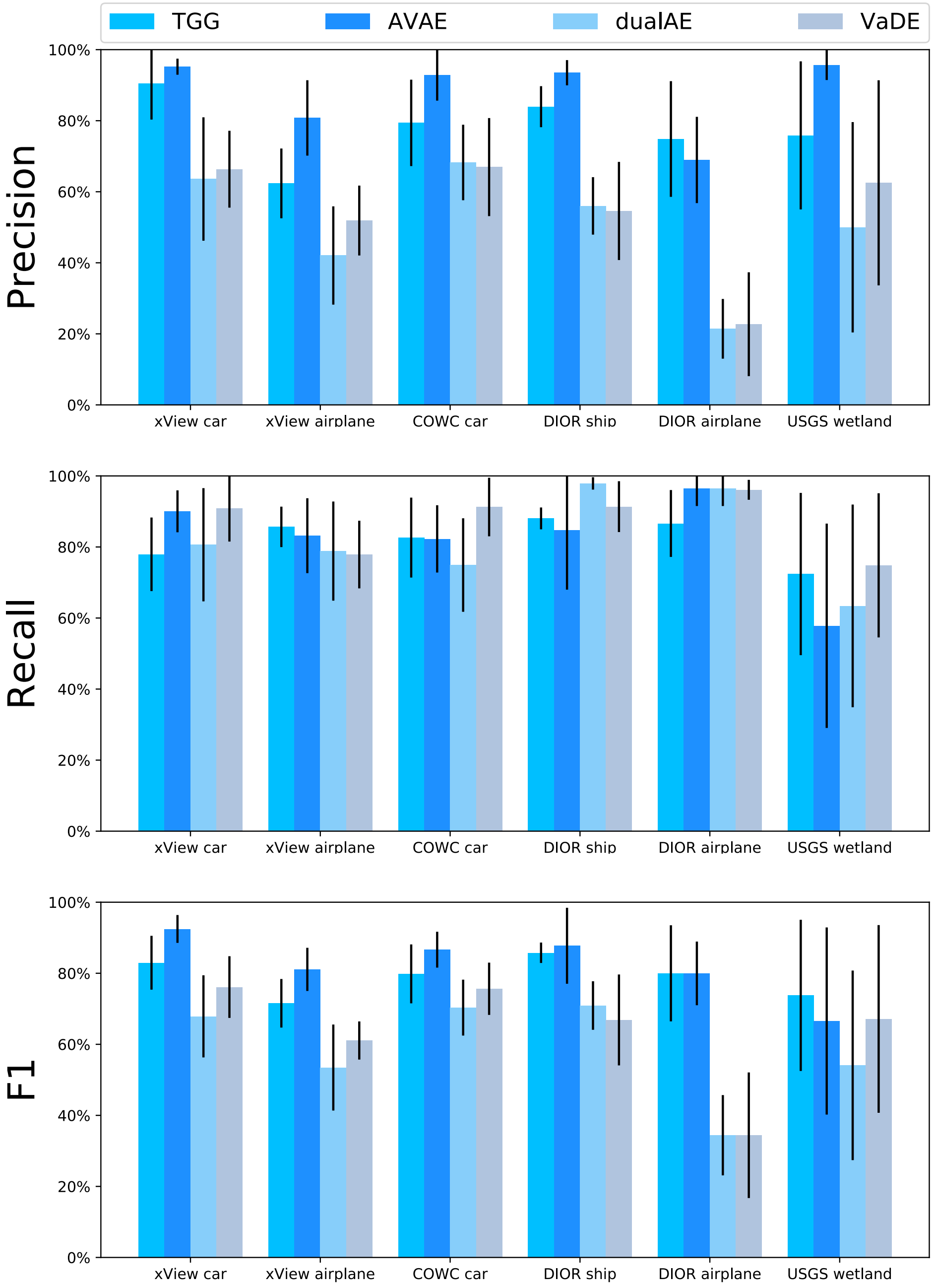
- TGGM outperformed the unsupervised generative clustering models, i.e., dualAE and VaDE
- TGGM's performance is similar to the semi-supervised generative model, i.e., AVAE



Cars in xView

True positive grid cells

False positive grid cells

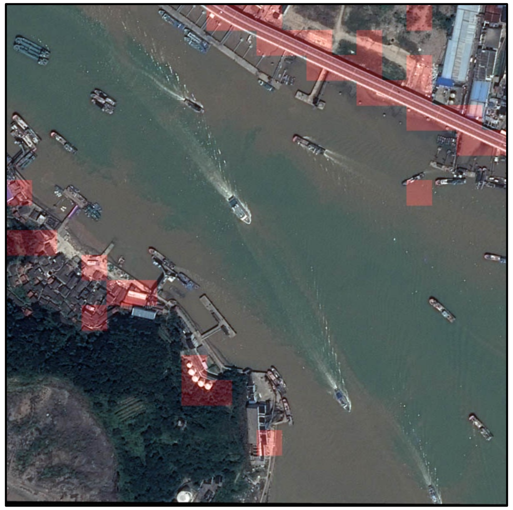


Experiment Results

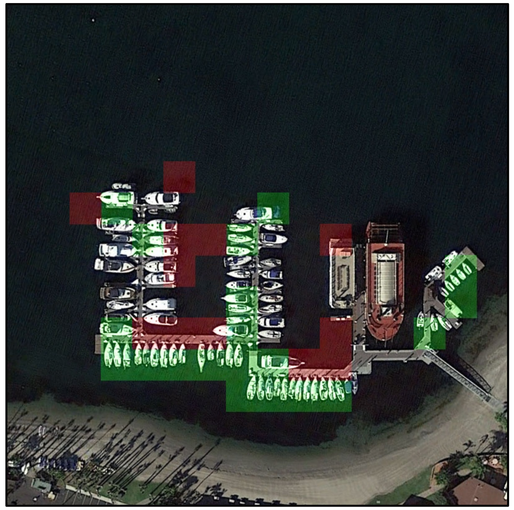
compared with supervised object detector, Yolov3

- For airplanes, TGGM: 60.15% mAP, while Yolov3: 72.2%
- For ships, TGGM: 69.92% mAP, while Yolov3: 87.4%
- Reasons for the low mAP
  - The weak semantic relationship between ROI and the target objects
  - Multi-scaled target objects

Airplanes



Ships



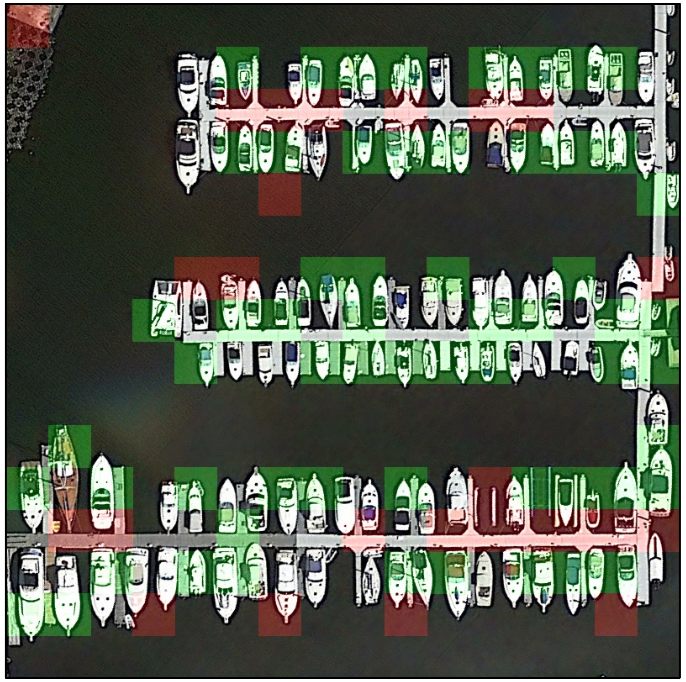
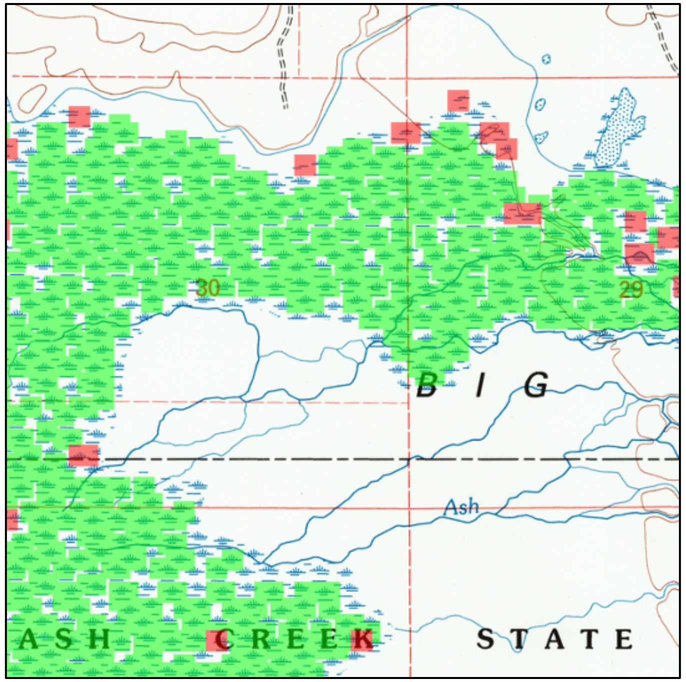
Ships

True positive grid cells

False positive grid cells

More Results Visualization

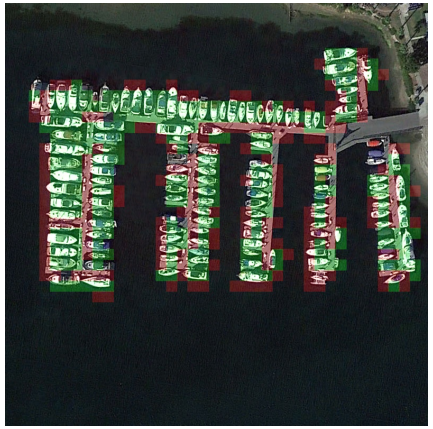
- When the ROI has strong semantic relationship with target objects
- The target objects' sizes are similar

Sensitivity Analysis

Grid-cell size for estimation

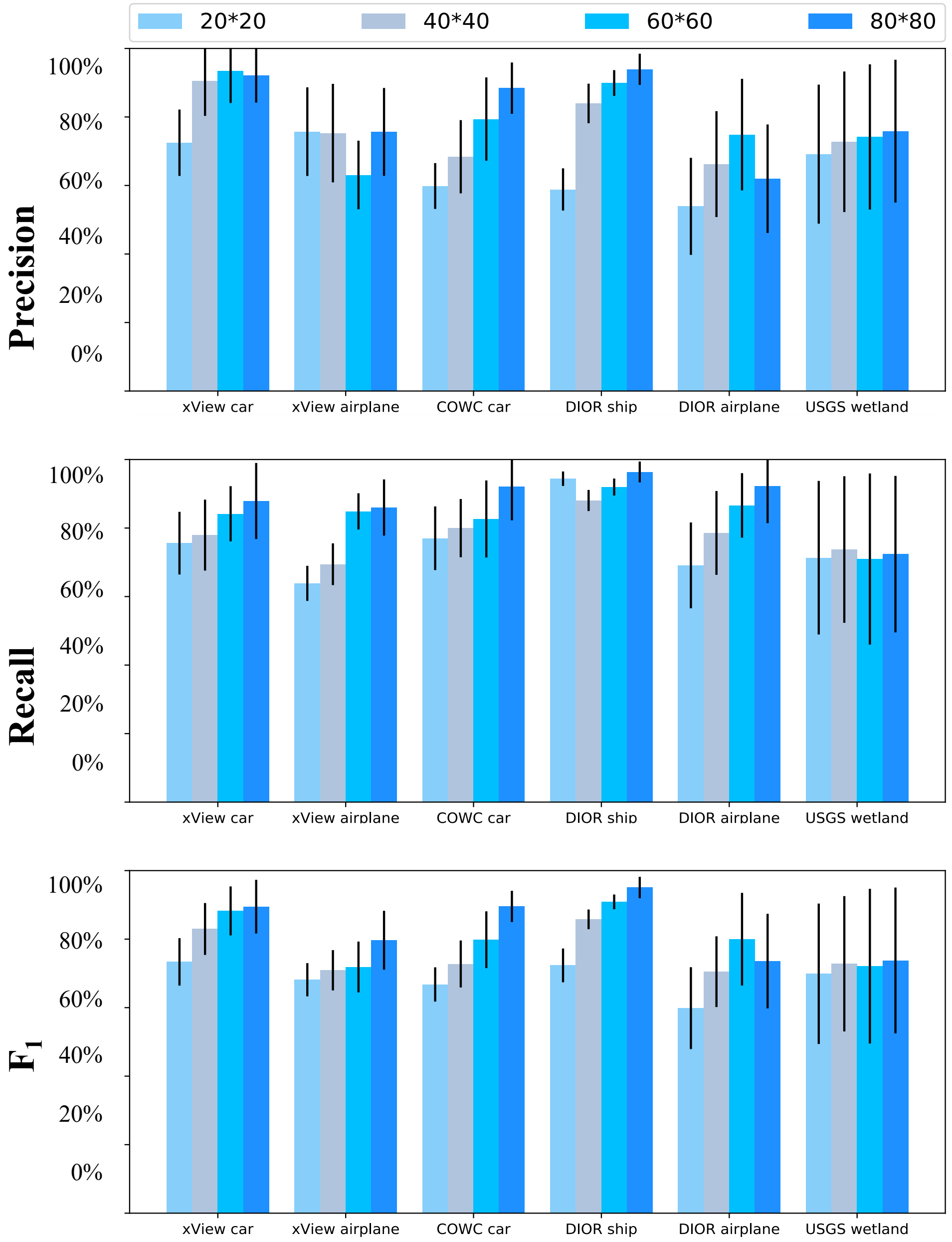
- Varied the grid-cell sizes
  - Grid-cell size precision, recall and F_1
  - Accuracy of spatial arrangement estimation



20-*20-pixel grid cell

40-*40-pixel grid cell

60-*60-pixel grid cell



Summary & Future Work

- TGGM estimate the **spatial arrangement** of target objects within **ROI**
- **Target-guidance mechanism** reduces the manual work to **one or a few labeled target objects**
- TGGM helps obtain **accurate** results
- Future work
  - apply to multi-scale target objects

Thank you!