# Data Synthesis For Object Recognition

Xi Zhang
Advisor: Professor Agam

xzhang22@hawk.iit.edu
Computer Science Department
Illinois Institute of Technology

# Overview

- Motivations and Importance of the problem.
- Introduction and novel contributions.
- Data synthesis in data space.
- Learning from synthetic data.
- Eliminating synthetic gap.
- Data synthesis in feature space.
- Conclusion

- Motivations and Importance of the problem.
- Introduction and novel contributions.
- Data synthesis in data space.
- Learning from synthetic data.
- Eliminating synthetic gap.
- Data synthesis in feature space.
- Conclusion.

# Motivation

1) To gain a good performance of a machine learning process, more high quality data is always desired.

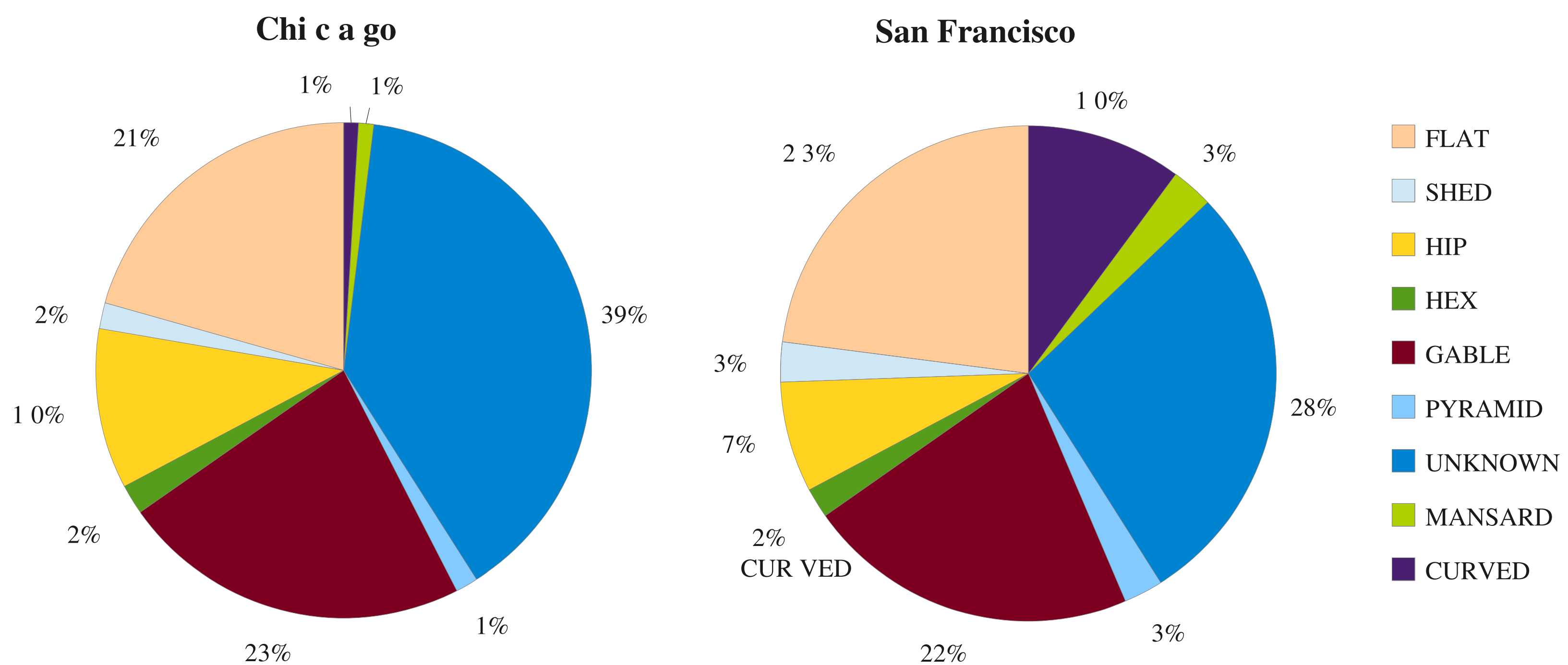
   a. Rare cases (absolute rarity)

   b. Rare classes (relative rarity)

1) To gain a good performance of a machine learning process, more high quality data is always desired.

a. Rare cases (absolute rarity)

b. Rare classes (relative rarity)



2) Supervised learning requires high quality labeled data

a. Time consuming, expensive.

b. Sometimes, impossible

2) Supervised learning requires high quality labeled data

a. Time consuming, expensive.

b. Sometimes, impossible

- Motivations and Importance of the problem.
- Introduction and novel contributions.
- Data synthesis in data space.
- Learning from synthetic data.
- Eliminating synthetic gap.
- Data synthesis in feature space.
- Conclusion.

# Introduction & Novel Contribution

- Solution: Data synthesis

- Challenges:

  1) Where to synthesize.

  2) How to synthesize.

- Novel contributions:

1)Data synthesis in data space.

2)Learning from synthetic data.

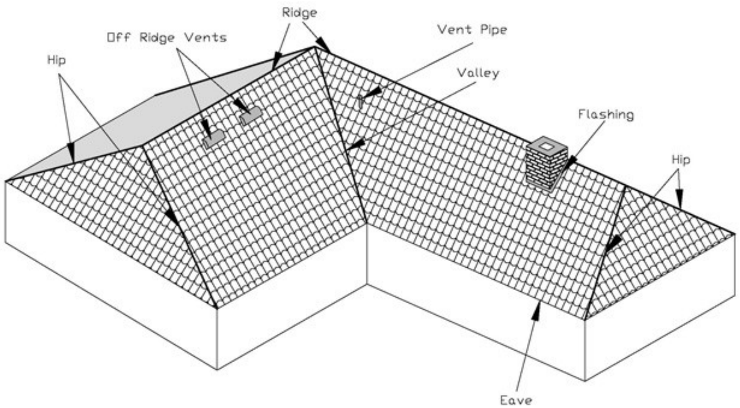
3)Eliminating synthetic gap.

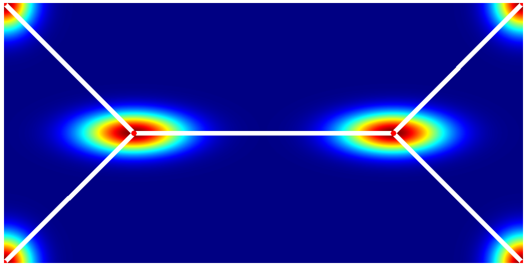
4)Data synthesis in feature space.

- Motivations and Importance of the problem.
- Introduction and novel contributions.
- Data synthesis in data space.
- Learning from synthetic data.
- Eliminating synthetic gap.
- Data synthesis in feature space.
- Conclusion.

# Data Synthesis in Data Space

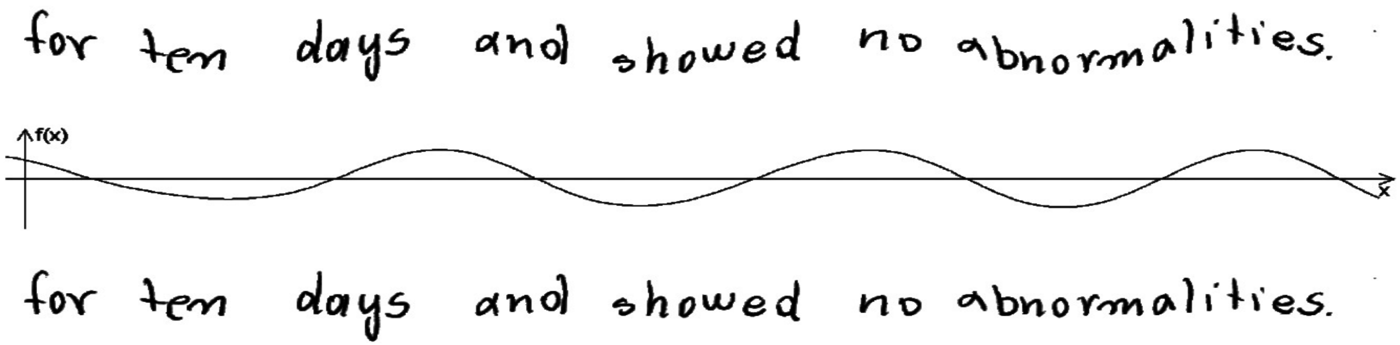
- Why in data space?
  - Most intuitive.
  - Sometime, easiest way.



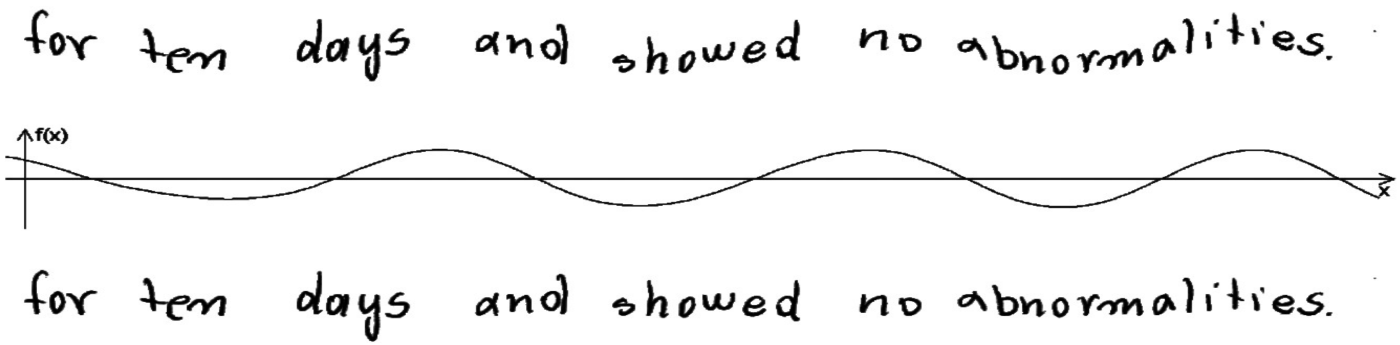



- But, some limitations too.
  - Require good prototype.
  - Application specific.

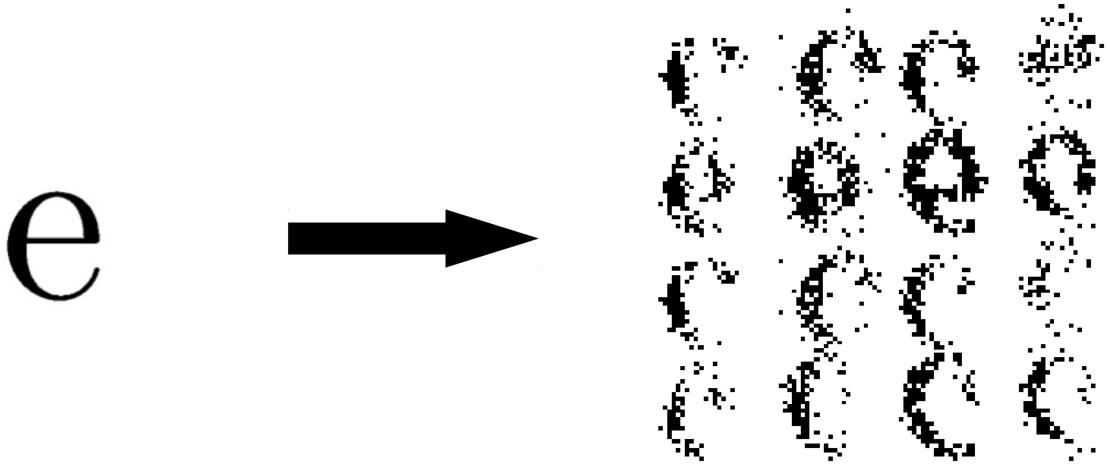
- Existing methods.
  - Geometric transformation. [89][90]



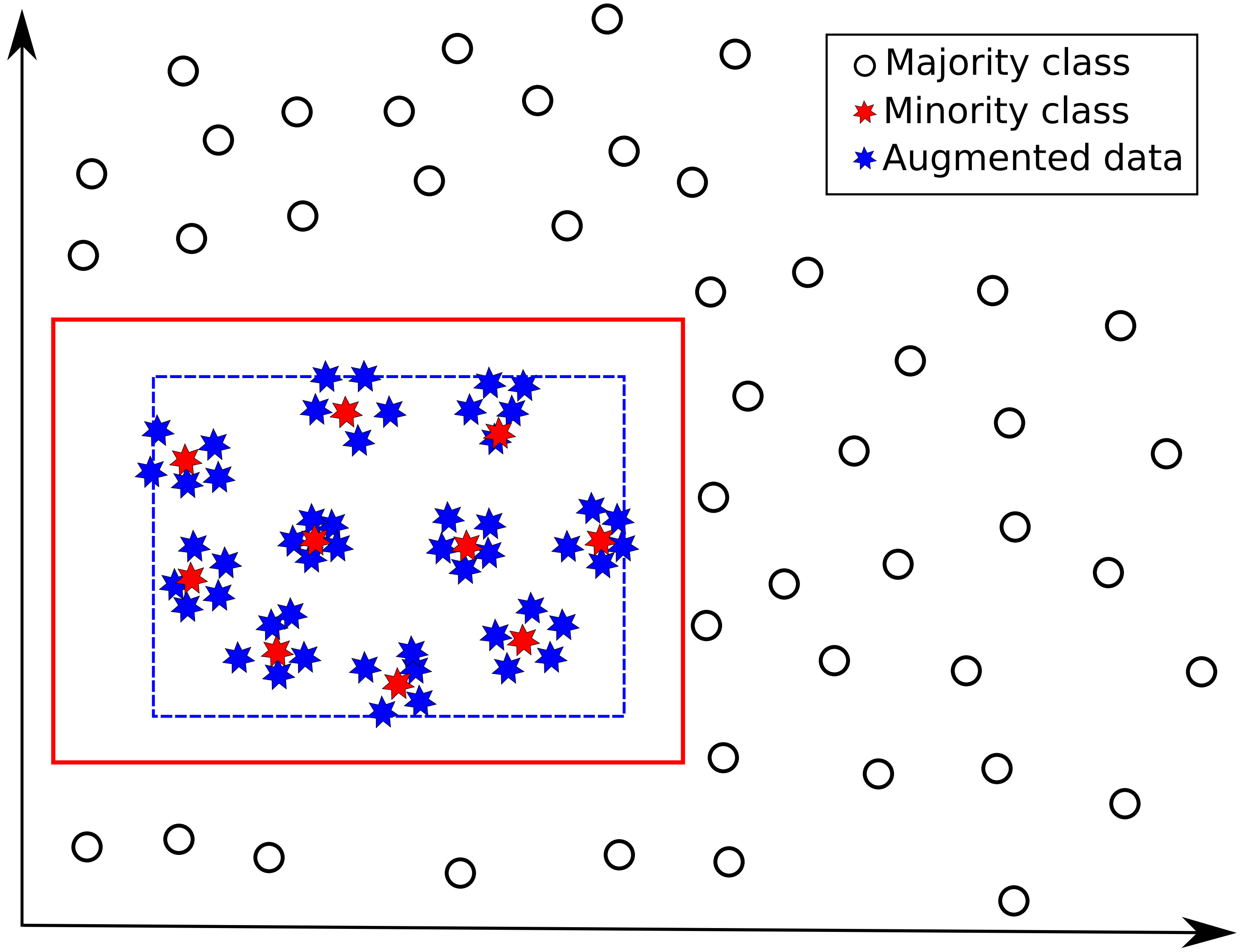
- Existing methods.
  - Geometric transformation. [89][90]




  - Image degradation. [4][65]



- Disadvantage of existing methods.

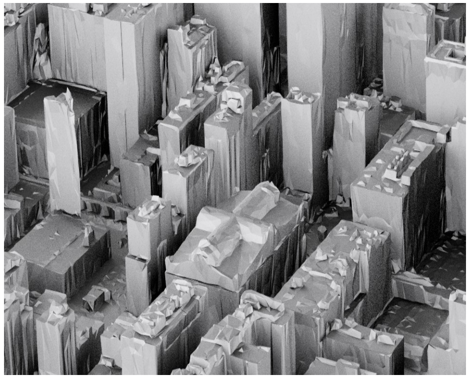


# The Proposed Approach

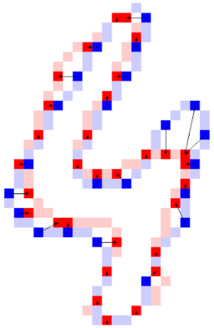
In contrast to existing methods, my approach:

- Generate templates/prototypes from data.
  - Automatically.
  - Semi-automatically.
- Synthesize more data using templates.

Satellite image roof style classification [100][105]



Aerial LiDAR roof style classification [106]



Handwritten digits recognition [105]

# Showcase One

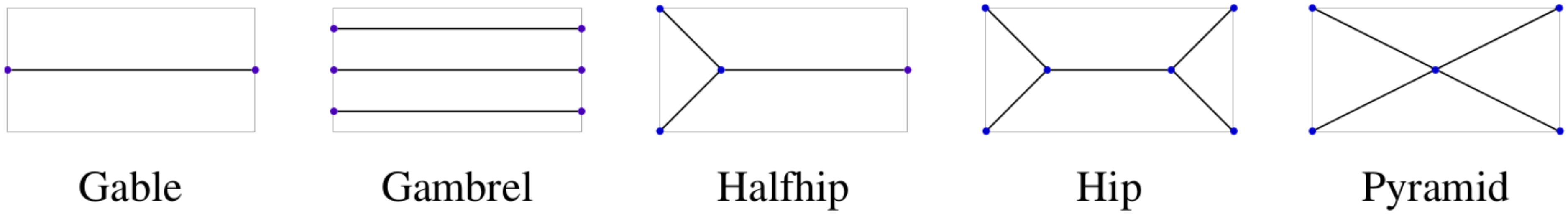
- Satellite image roof style classification.

  Decompose roof if necessary.

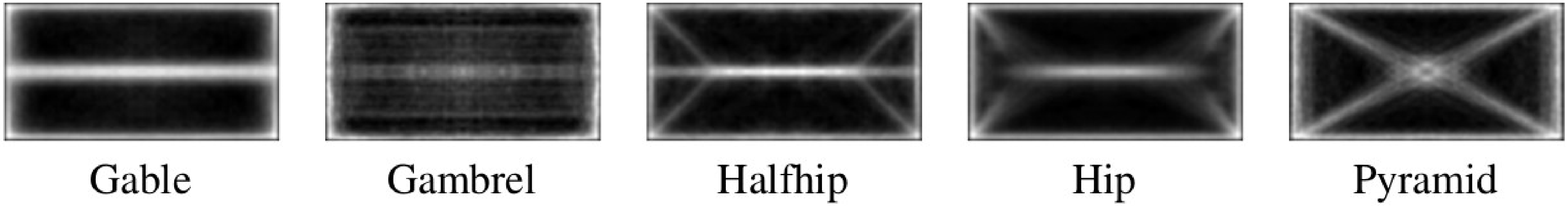
- Preprocessing to refine, scale and align images.

## Showcase One

- Prototype by design.
  - In [100], I manually designed prototypes.

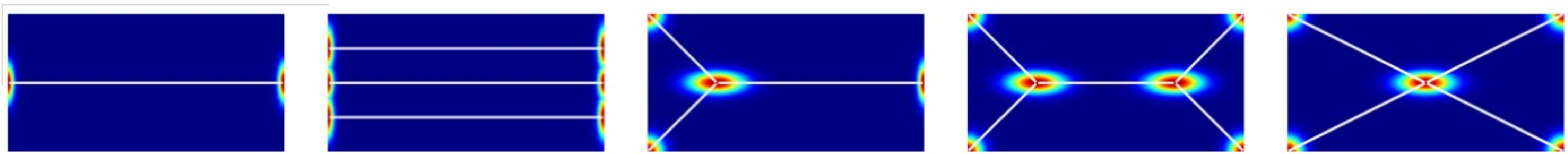



- Prototype by computation.
  - Identification of control points done automatically.



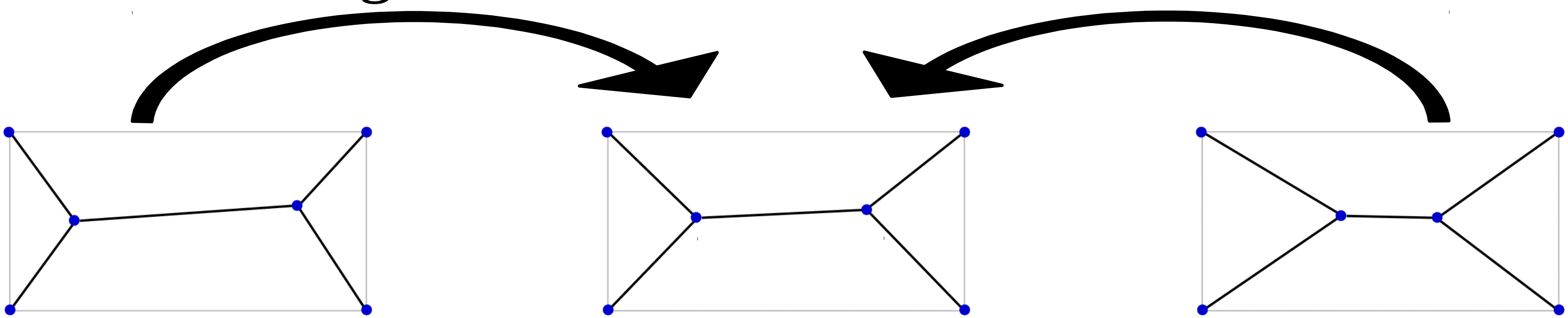
# Showcase One

- Data synthesis by distribution of control points.



Gable　Gambrel　Halfhip　Hip　Pyramid

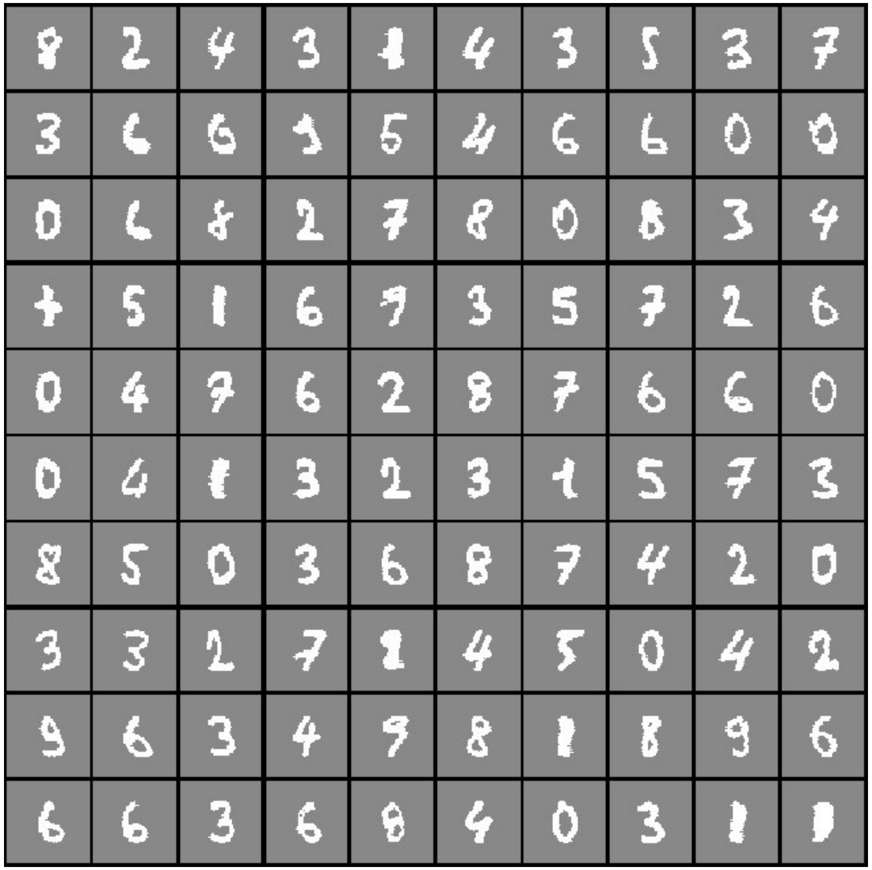
- Data synthesis by interpolation/extrapolation between nearest neighbors.



Source 1　Synthetic　Source 2

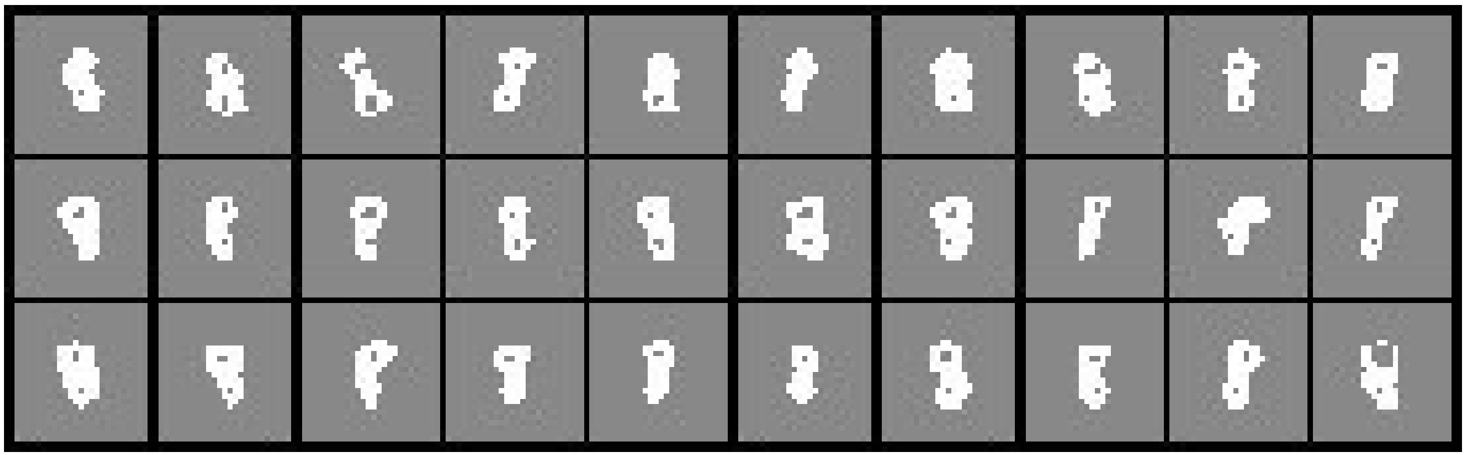
# Showcase Two

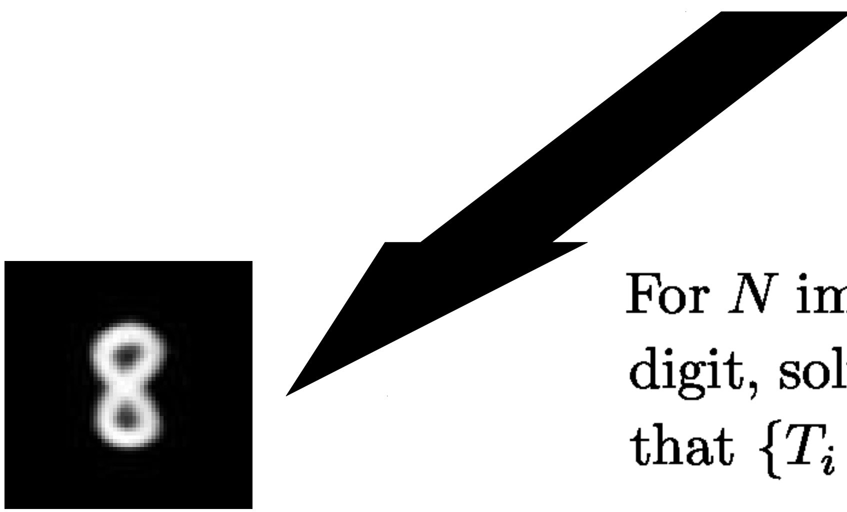
- Absolute rarity, digit from 0 ~ 9, each has around 200 samples.



# Showcase Two

- Generate prototype by congealing. [61]

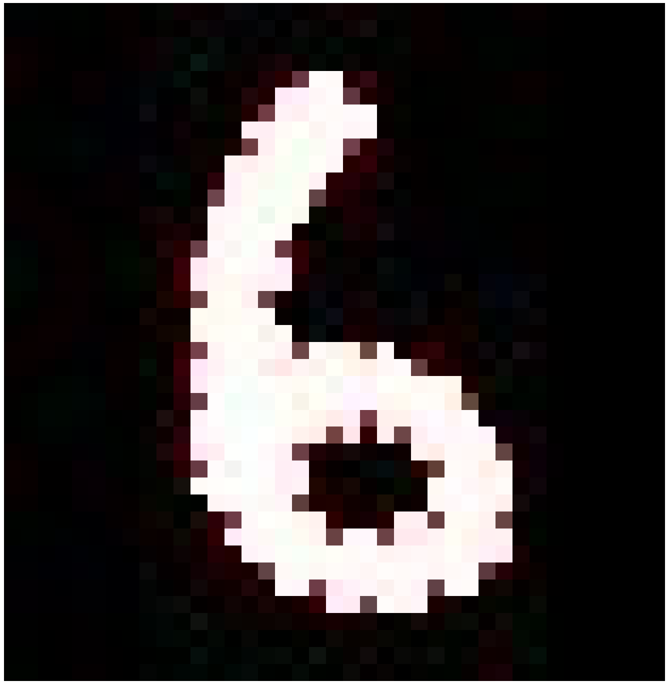




For $N$ images $\{I_i\}_{i=1}^{N}$ with the same digit, solve for transformations $\{T_i\}_{i=1}^{N}$, that $\{T_i \cdot I_i\}_{i=1}^{N}$ minimize joint entropy

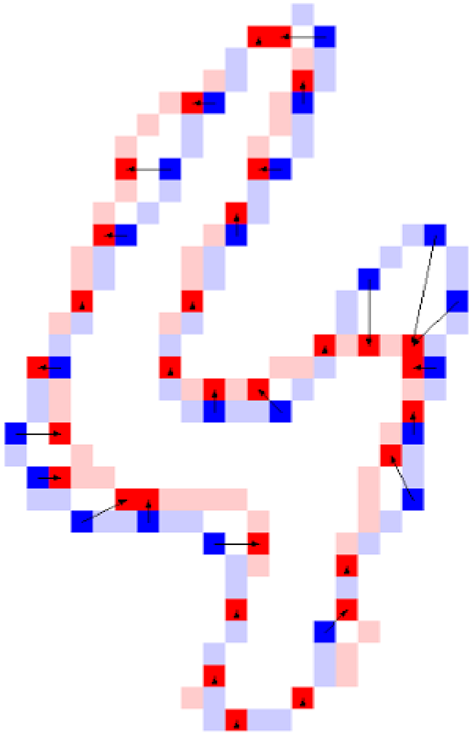
# Showcase Two

- Building correspondence among data.
  - Set control points on prototype.



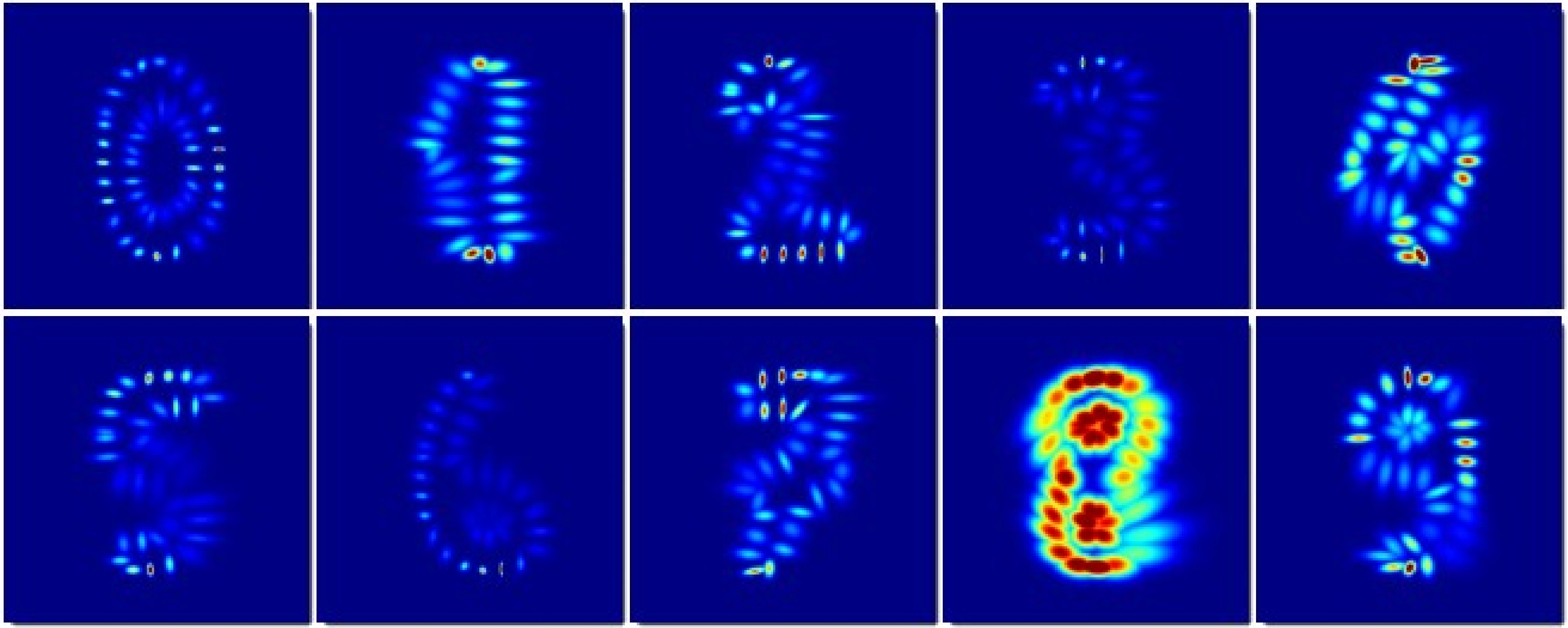
# Showcase Two

- Building correspondence among data.
  - Set control points on prototype.
  - Find corresponding control points on each data.



# Showcase Two

- Data synthesis by distribution of control points.



- Data synthesis by interpolation/extrapolation between nearest neighbors.

# Showcase Two

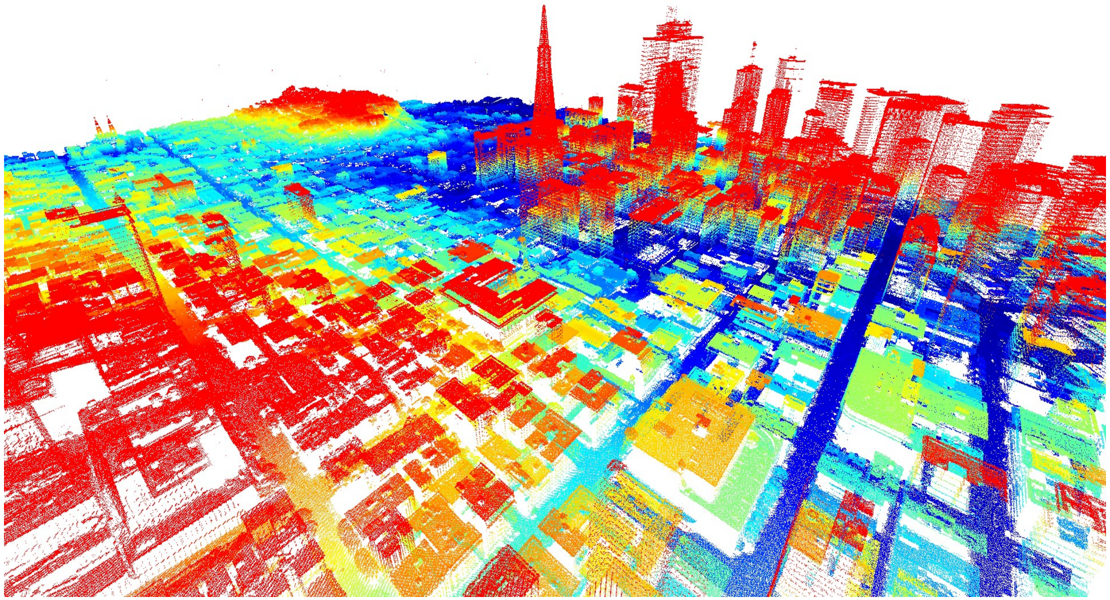
- Boost the performance by adding synthetic data.

| | Real | Syn | Real+Syn |
|---|---|---|---|
| **CNN** | 0.65 | 0.68 | 0.70 |
| **SVM** | 0.77 | 0.78 | 0.80 |

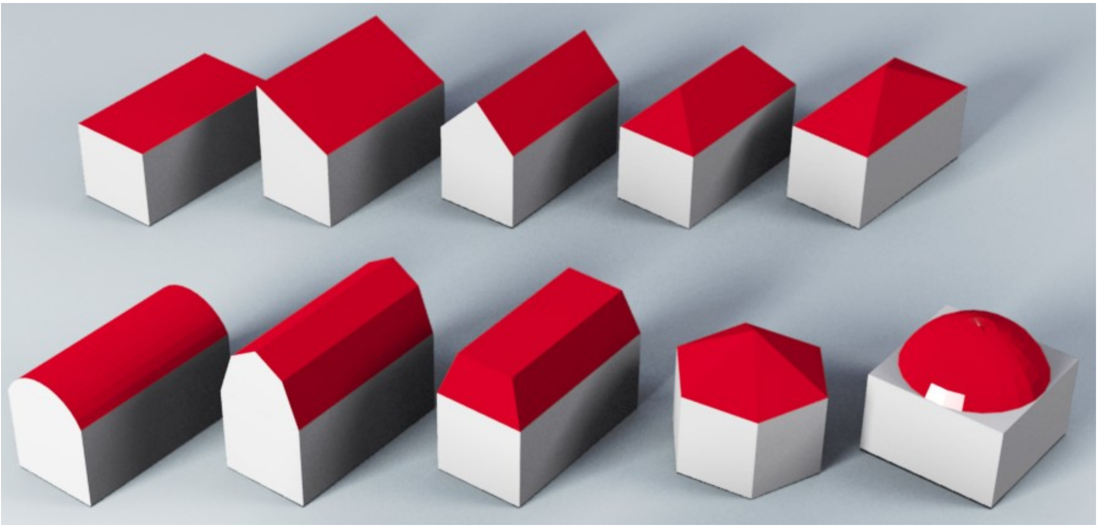
# Showcase Three

Aerial LiDAR

Downtown San Francisco

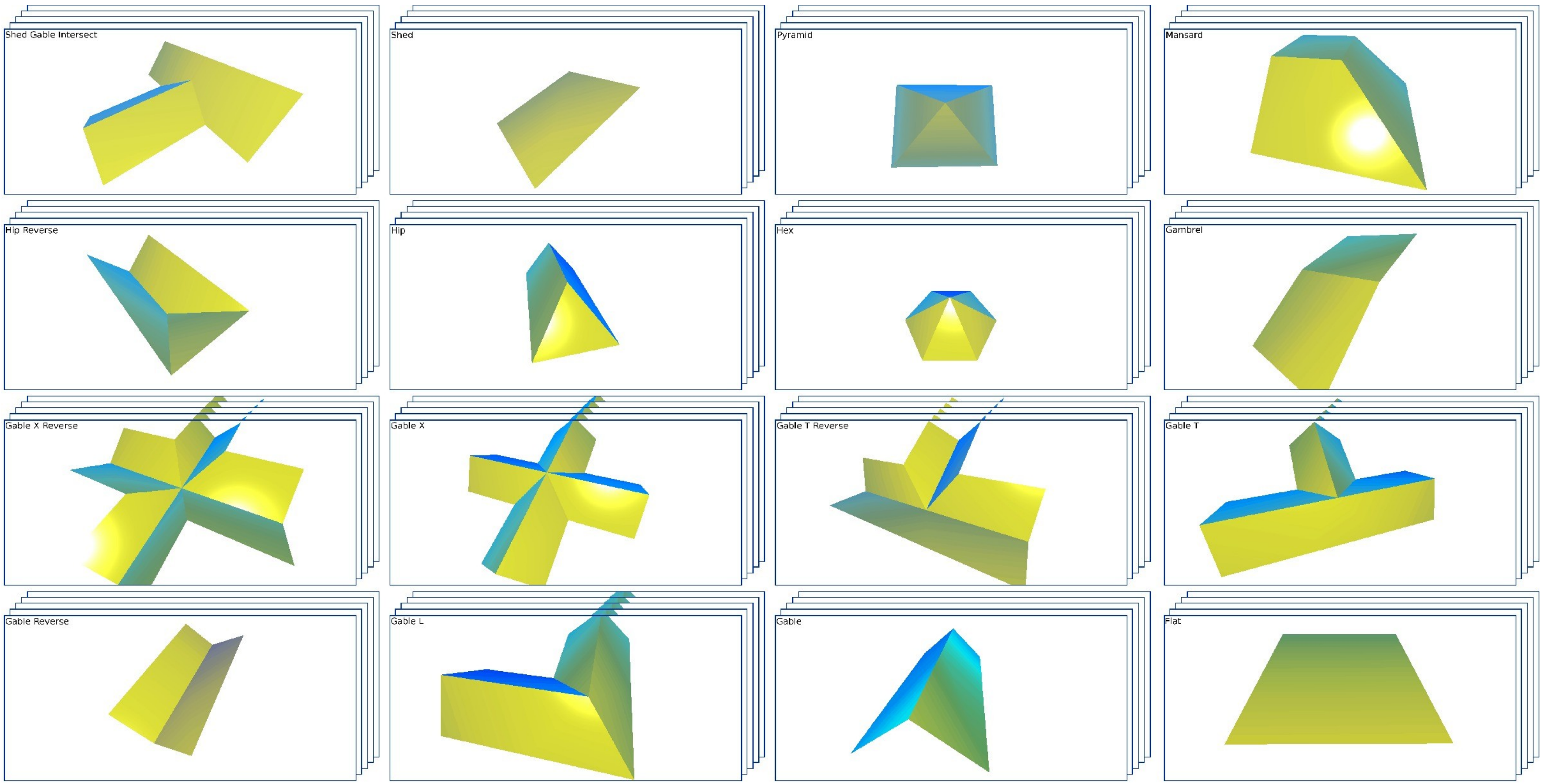


Candidate roof styles



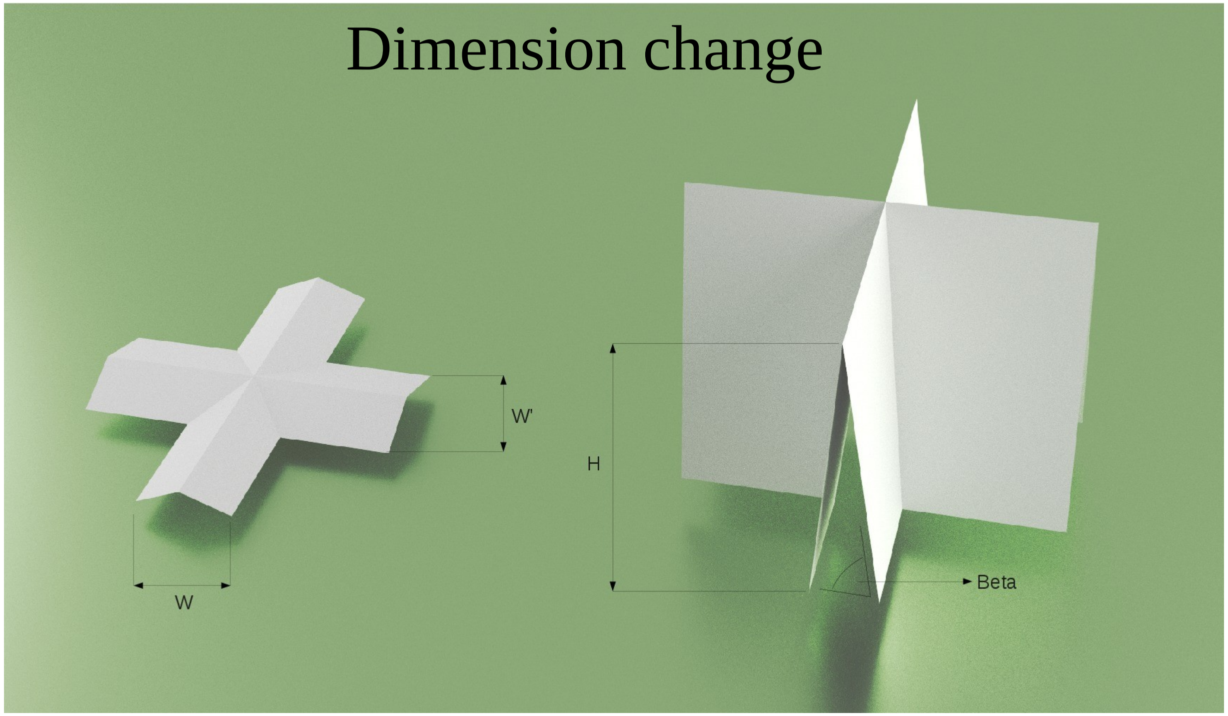
# Showcase Three

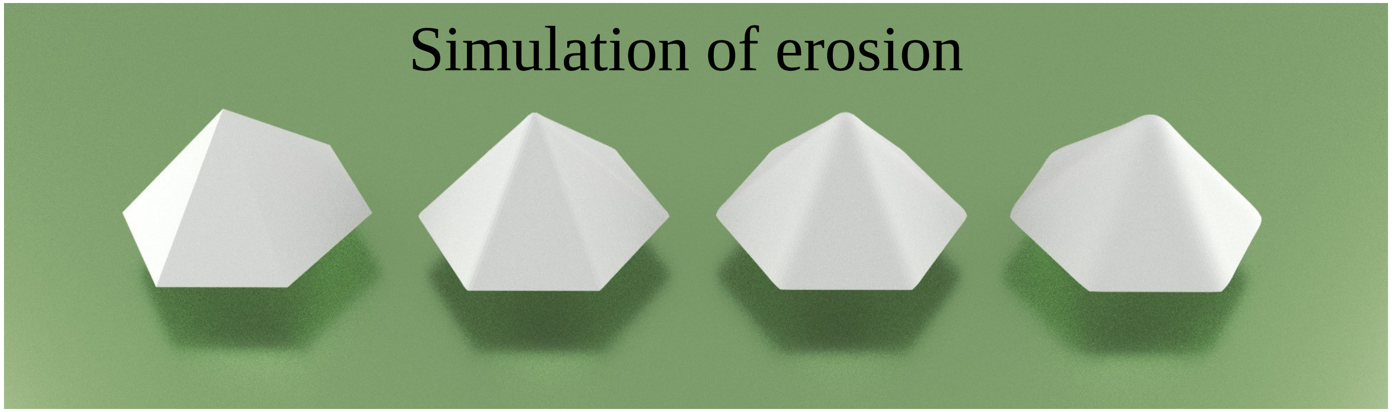
- Build prototypes.



# Showcase Three

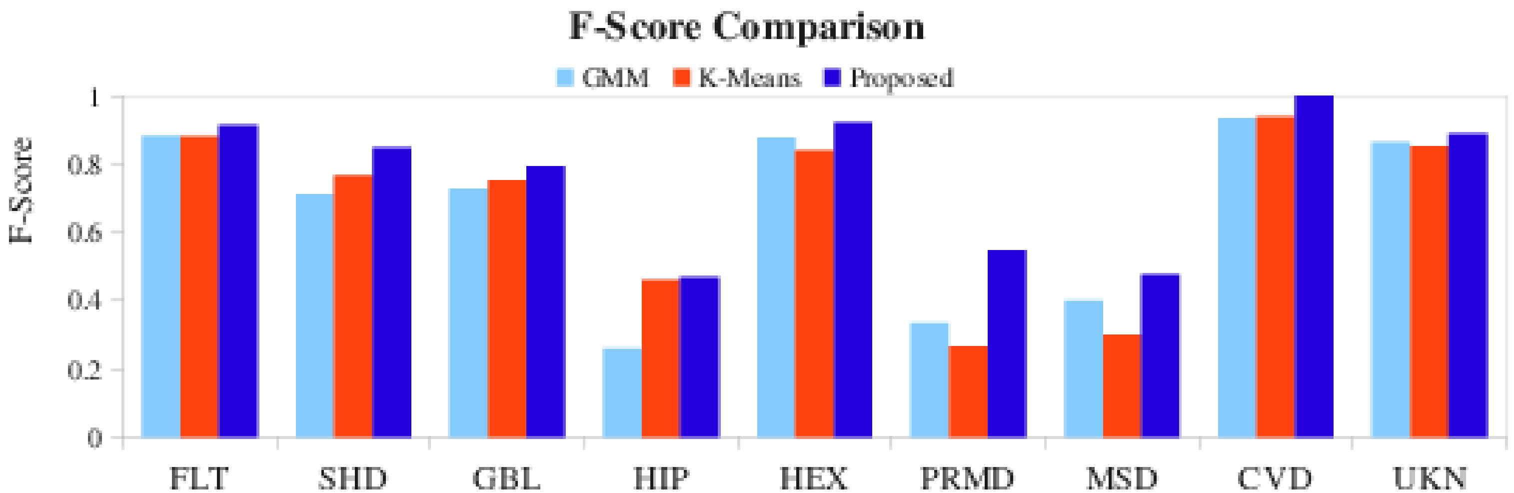
- Derive more data.



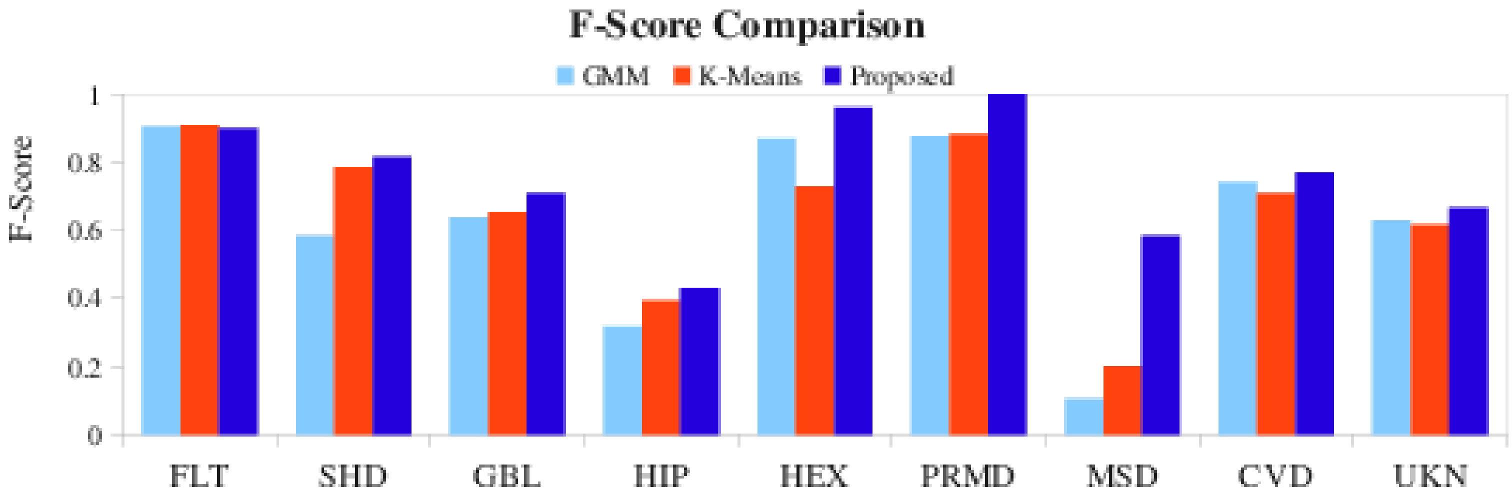



# Showcase Three

- Boost recognition rate by adding synthetic data.




Chicago

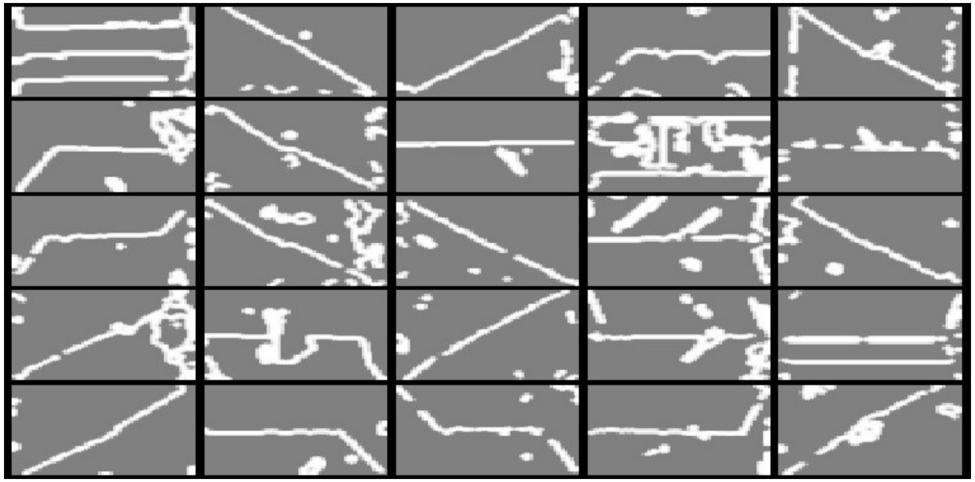



San Francisco

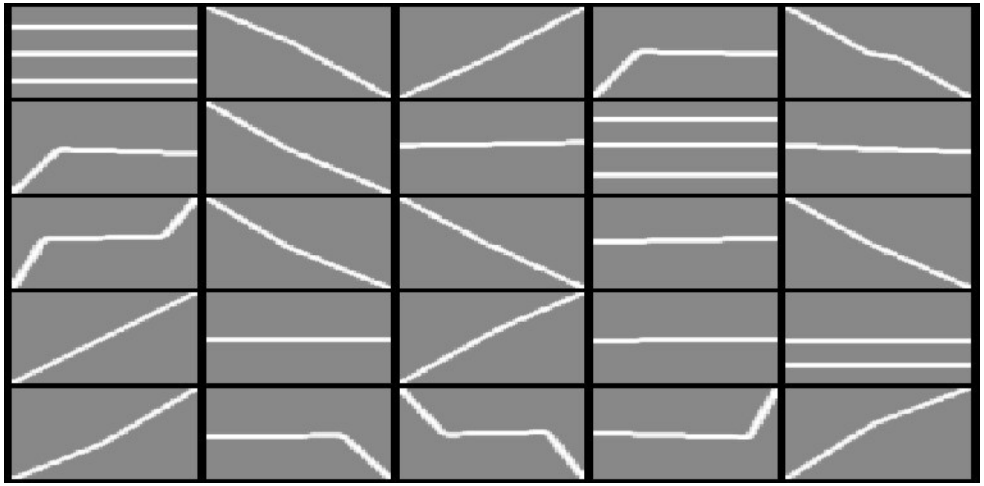
- Motivations and Importance of the problem.
- Introduction and novel contributions.
- Data synthesis in data space.
- Learning from synthetic data.
- Eliminating synthetic gap.
- Data synthesis in feature space.
- Conclusion.

# Learning from Synthetic Data

- Challenging task.

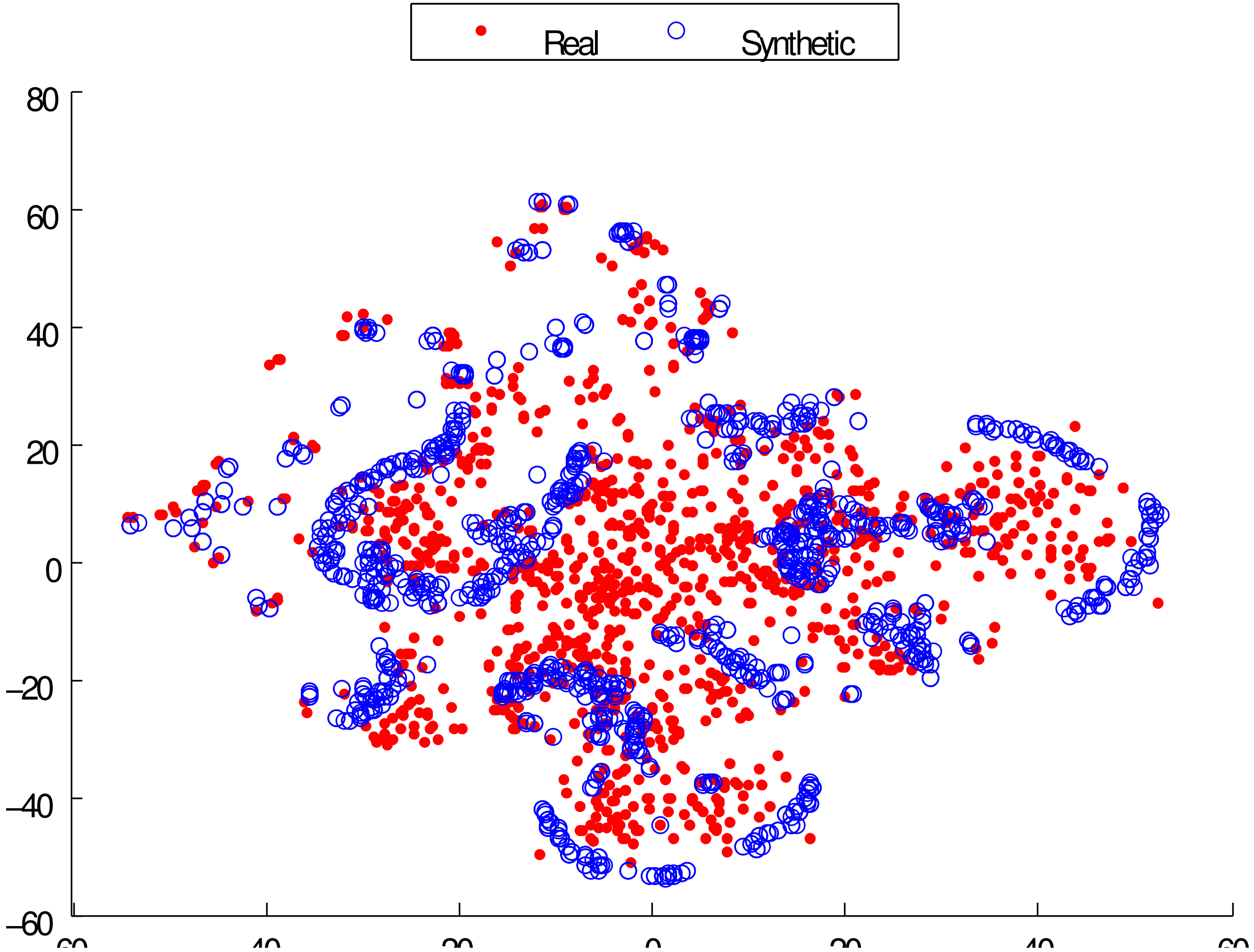


Actual data



Synthetic data

- Challenging task.



- Previous work.

  They all treat synthetic data as actual ones.[39][87][66][89][90][65]

- The proposed approaches.

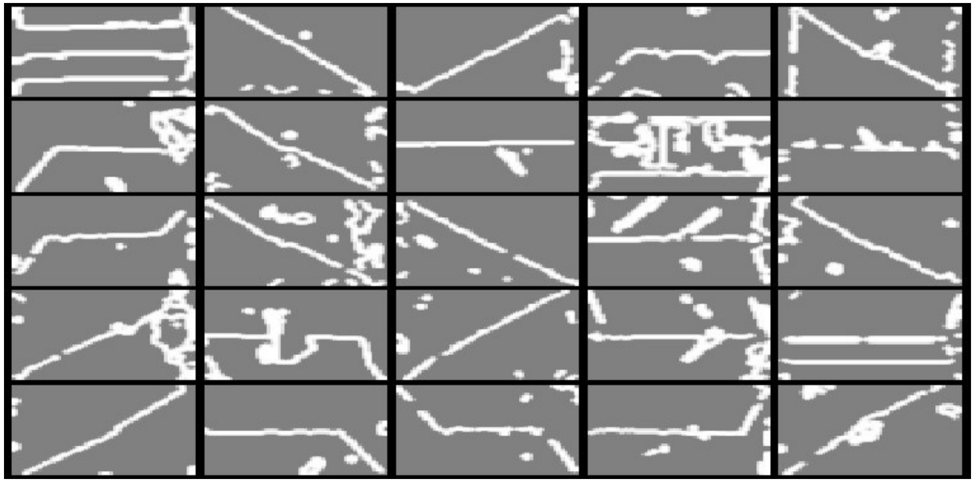
  I build features that contain equivalent amount of information between synthetic data and actual data.
- Two types of features
  - Type I: Ignore additional information in actual data.
  - Type II: Compensate additional information for synthetic data.

Two types of features:

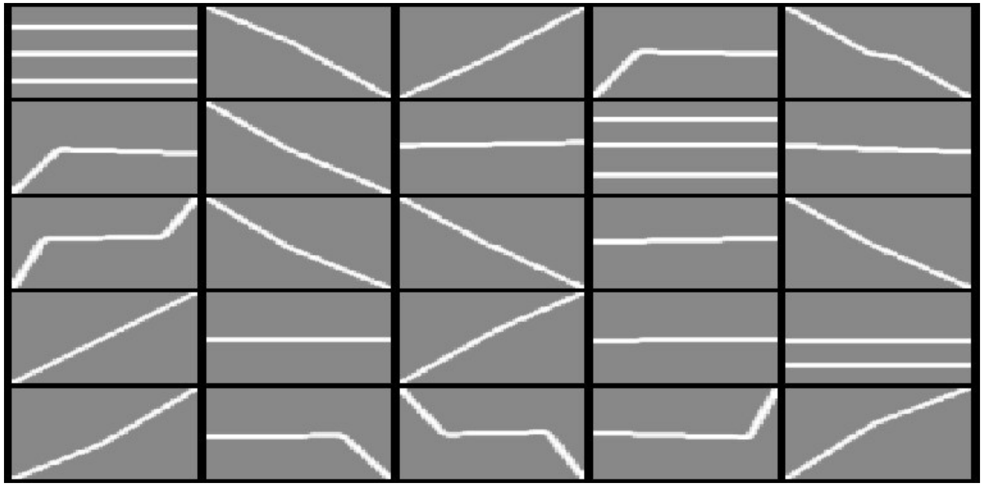
- Type I: Ignore additional information in actual data.
- Type II: Compensate additional information for synthetic data.

# Showcase One

- Satellite image roof style classification.



Actual data

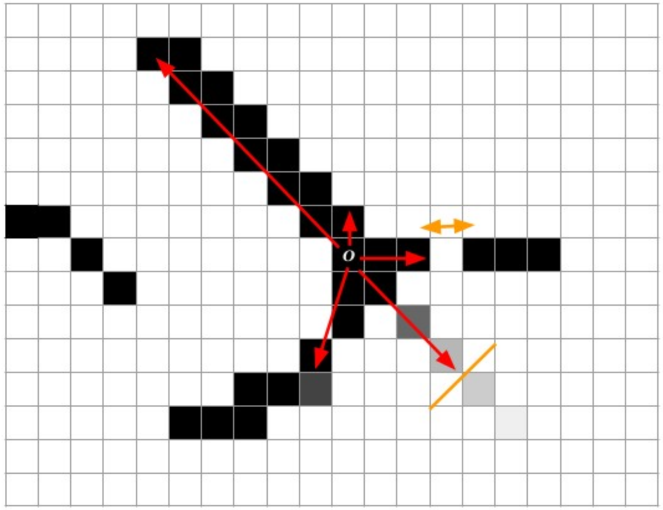


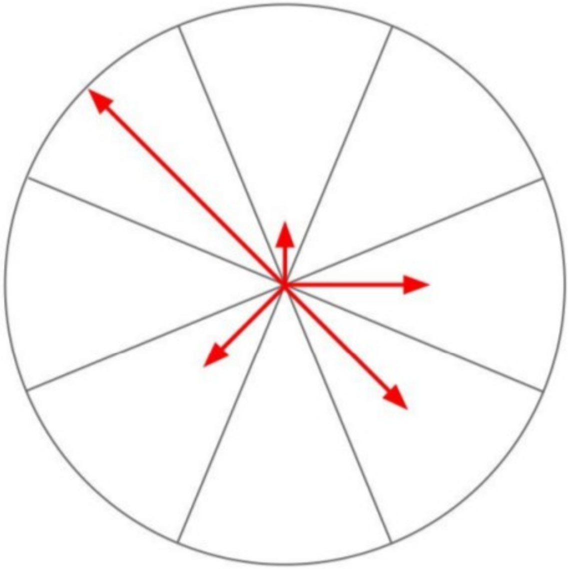
Synthetic data

- Features expected:
  - Ignore small blobs in actual data.
  - Highlight the most evident structure of roofs.

# Showcase One

- We proposed a feature called Histogram of Ray (HOR) in [100]
  - Highlight edge length and direction in [100].
  - Translation, scale, rotation invariant.



## Showcase One

- Compare with several well-known image features:

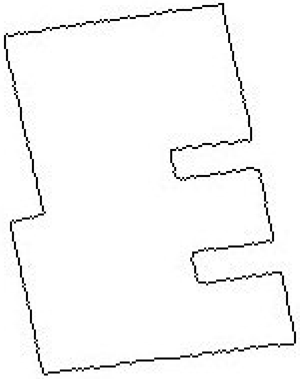
| | HIP | GABLE | FLAT | HALFHIP |
|---|---|---|---|---|
| HOG | 0.805 | 0.882 | 0.954 | 0.597 |
| SC | 0.350 | 0.828 | 0.959 | 0.140 |
| HOR | 0.898 | 0.950 | 0.968 | 0.632 |
| LBP | 0.000 | 0.986 | 0.631 | 0.000 |
| HOR+HOG | 0.931 | 0.959 | 0.982 | 0.667 |
| HOR+SC | 0.619 | 0.891 | 0.959 | 0.436 |
| HOG+SC | 0.752 | 0.959 | 0.945 | 0.474 |
| SC | 0.743 | 0.869 | 0.963 | 0.509 |

# Showcase Two

- Retrieve the most matching satellite building image for synthetic building footprint. [104]
- Challenges: Noise, occlusion and shape difference.

Original actual image

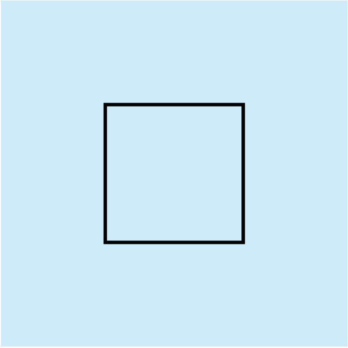
Edge extracted

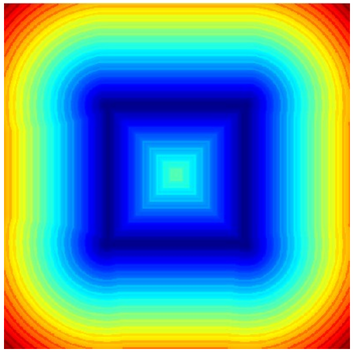


Synthetic building footprint

# Showcase Two

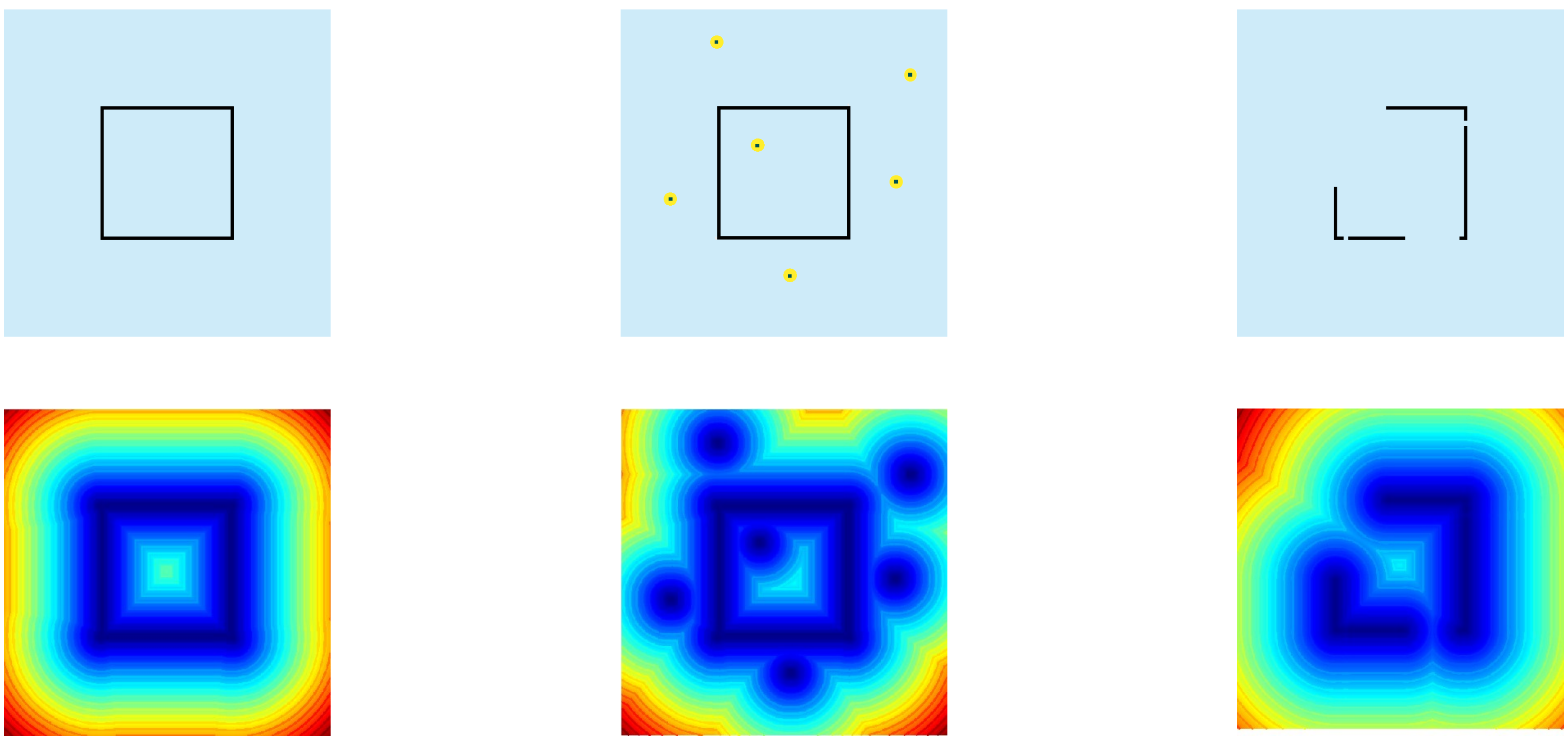
- Matching metric is derived from classic Chamfer matching.





# Showcase Two

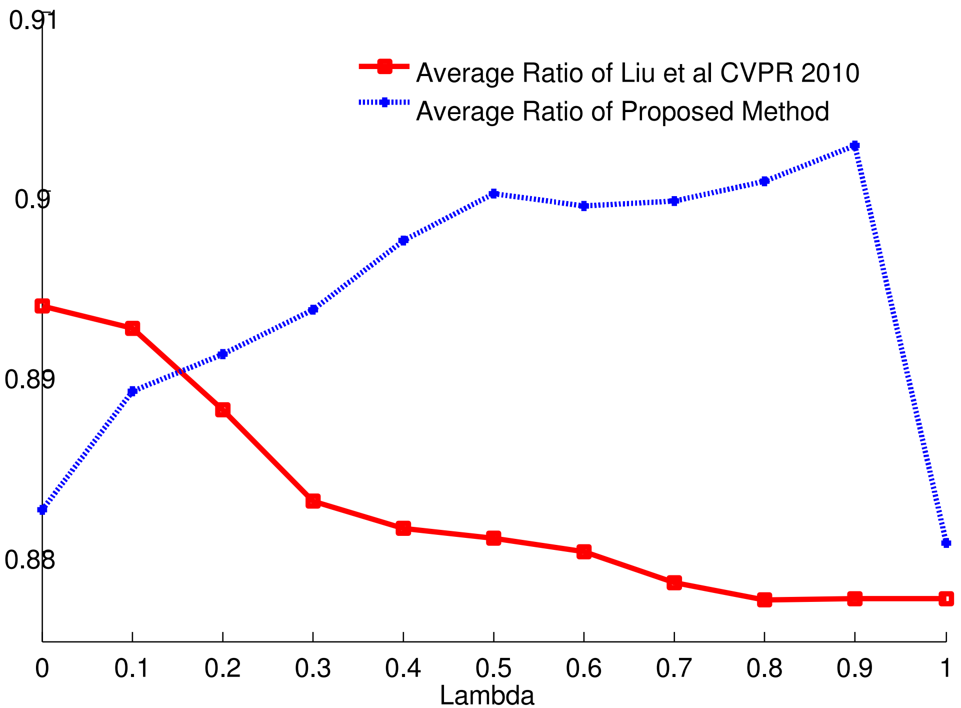
- Matching metric is derived from classic Chamfer matching.
- However



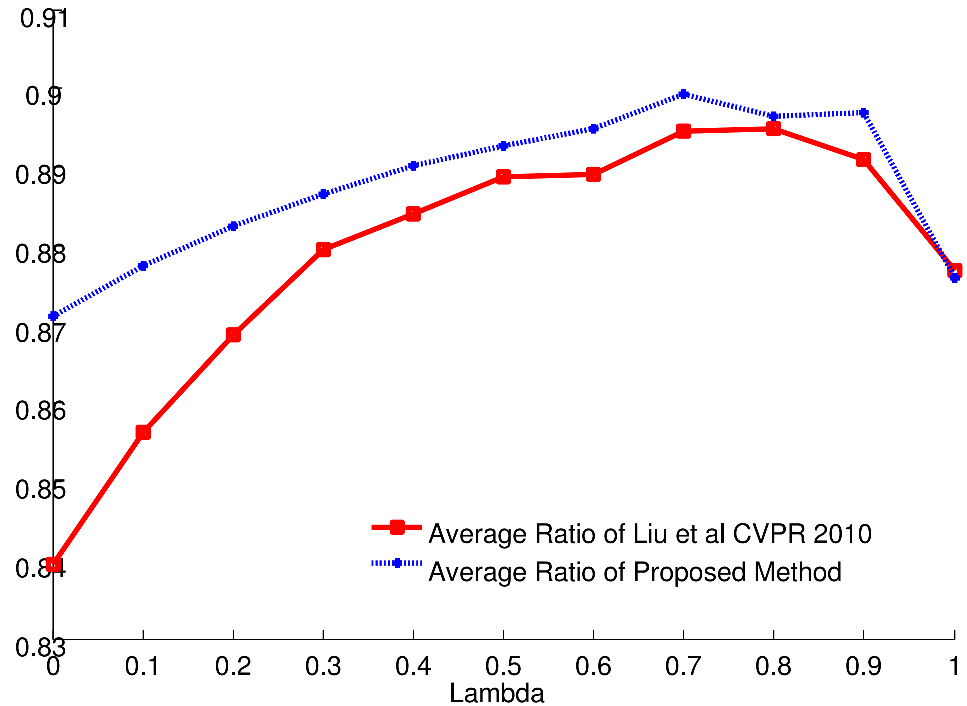
## Showcase Two

- I extend Chamfer matching from three aspects.
  - Add edge orientation to distance metric.
  - Add edge distance variance to distance metric.
  - Mechanism to matching incomplete targets.

# Showcase Two



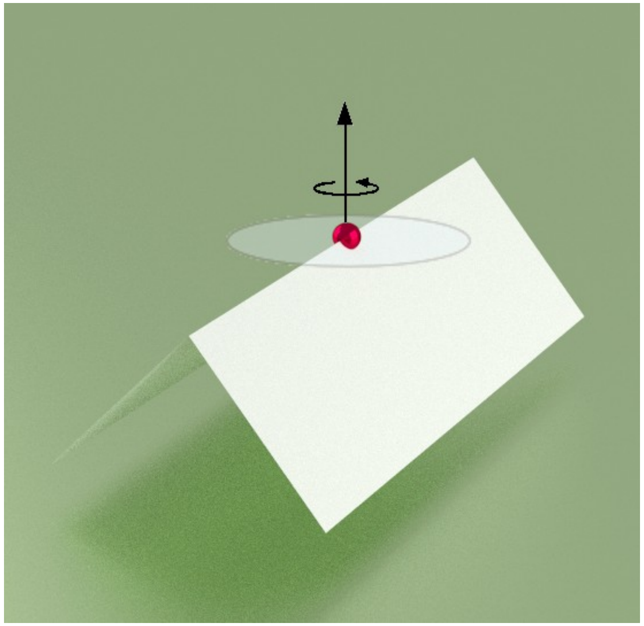

Chicago

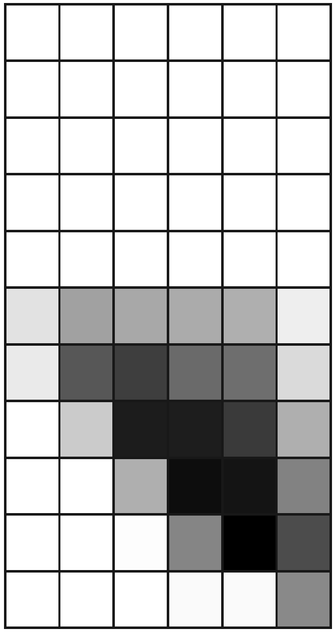



San Francisco

# Showcase Three

- Extract features that characterize local geometry.





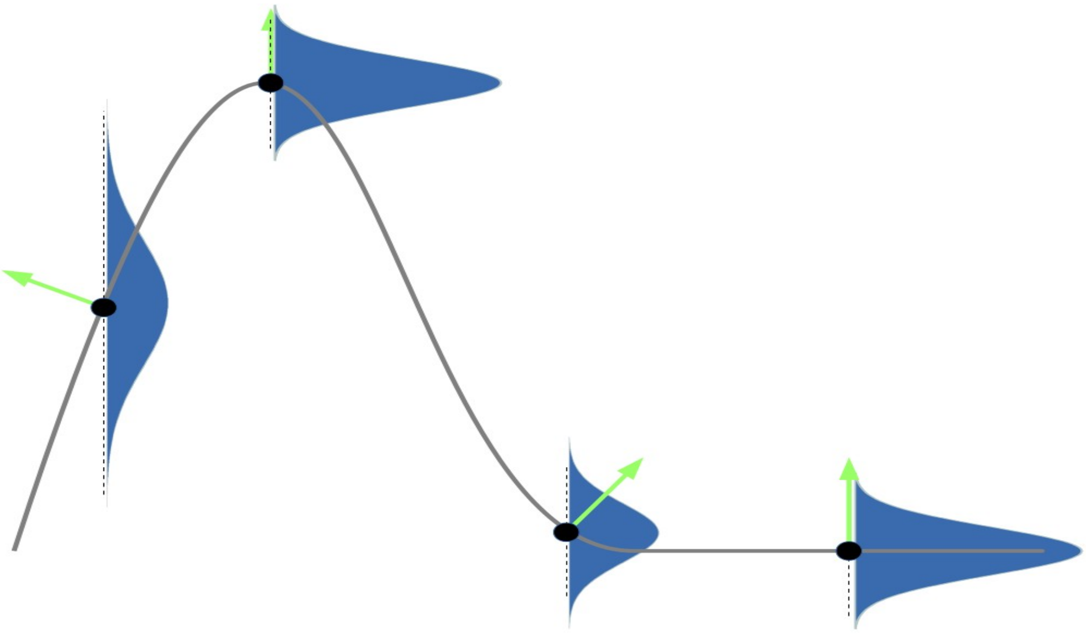
Spin image

# Showcase Three

- Extract features that characterize local geometry.
- However, too regular and smooth.

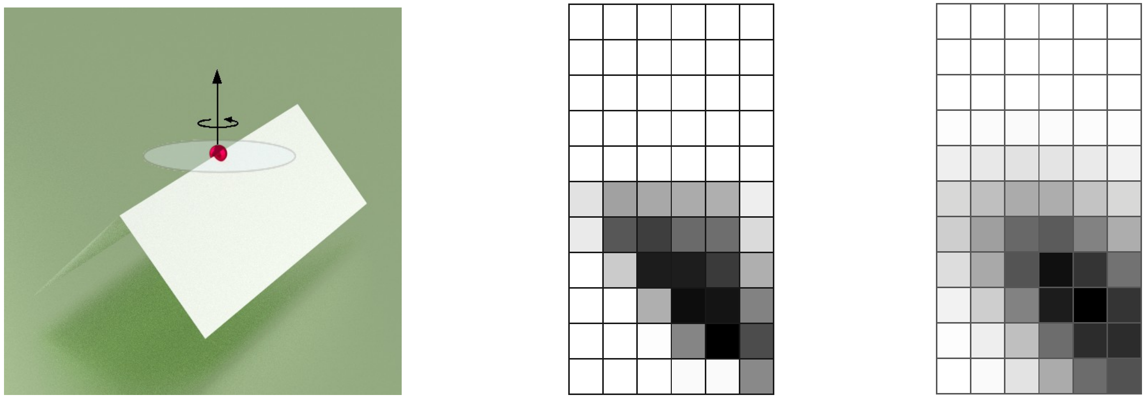
# Showcase Three

- Extract features that characterize local geometry.
- However, too regular and smooth.
- Learning bumpiness as a function of surface slope.



## Showcase Three

- Extract features that characterize local geometry.
- However, too regular and smooth.
- Learning bumpiness as a function of surface slope.
- Add random noise to synthetic data using knowledge learned.



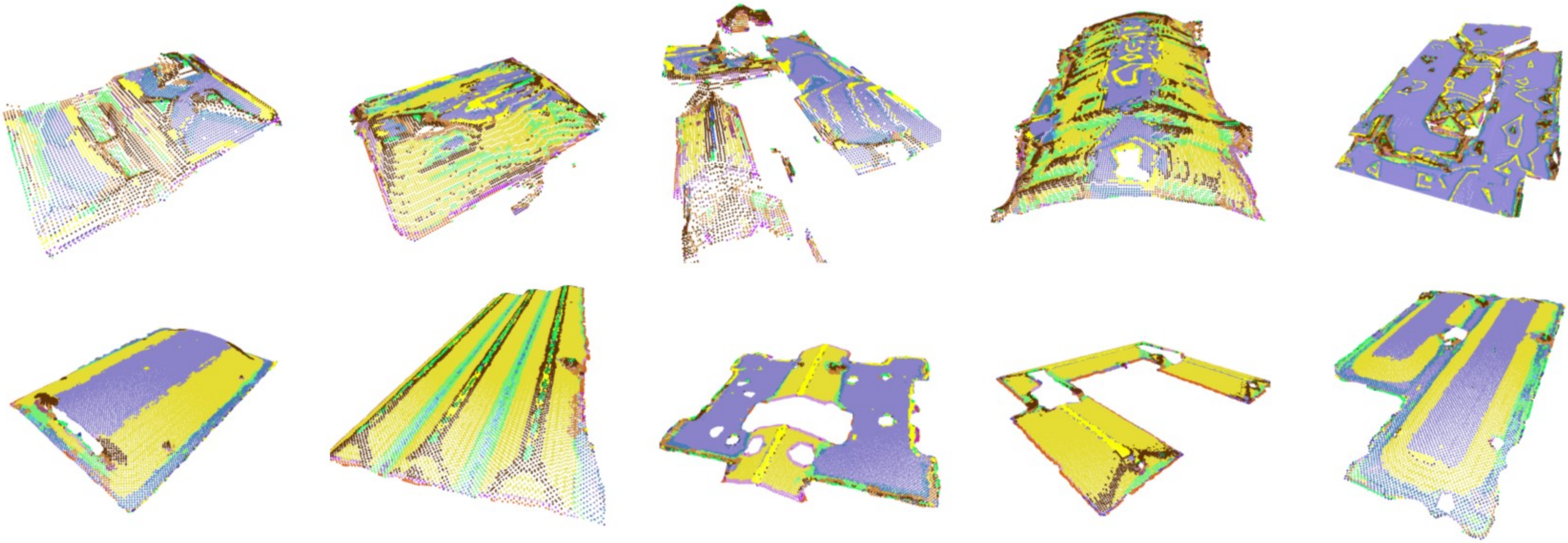
Spin image

# Showcase Three


ILLINOIS INSTITUTE OF TECHNOLOGY

## Showcase Three

- Point semantics classification results.



# Showcase Three

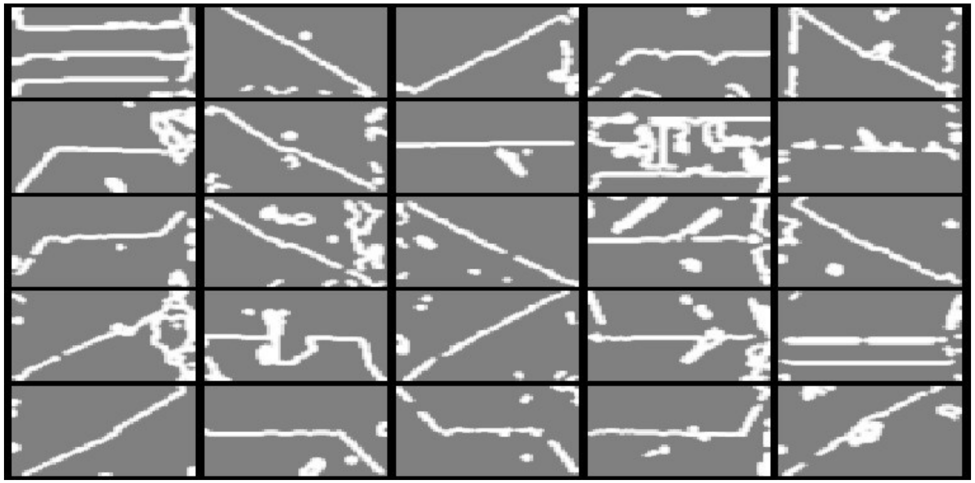
- Roof style classification results compared to unsupervised approaches.

| | Chicago | | | | | | San Francisco | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Precision | | | Recall | | | Precision | | | Recall | | |
| | GMM | KM | Ours | GMM | KM | Ours | GMM | KM | Ours | GMM | KM | Ours |
| FLAT | 0.87 | 0.85 | 0.92 | 0.88 | 0.90 | 0.90 | 0.88 | 0.88 | 0.86 | 0.93 | 0.93 | 0.93 |
| SHED | 0.94 | 0.94 | 0.91 | 0.57 | 0.64 | 0.78 | 0.87 | 0.91 | 1.00 | 0.43 | 0.68 | 0.68 |
| GABLE | 0.62 | 0.67 | 0.71 | 0.86 | 0.84 | 0.88 | 0.57 | 0.61 | 0.65 | 0.71 | 0.69 | 0.77 |
| HIP | 0.65 | 0.63 | 0.63 | 0.16 | 0.36 | 0.37 | 0.55 | 0.61 | 0.70 | 0.22 | 0.28 | 0.31 |
| HEX | 0.87 | 0.86 | 0.93 | 0.87 | 0.81 | 0.90 | 0.90 | 0.80 | 0.92 | 0.83 | 0.66 | 1.00 |
| PYRAMID | 0.83 | 0.66 | 1.00 | 0.20 | 0.16 | 0.37 | 0.87 | 0.83 | 1.00 | 0.87 | 0.93 | 1.00 |
| MANSARD | 1.00 | 0.75 | 1.00 | 0.25 | 0.18 | 0.31 | 0.50 | 0.66 | 1.00 | 0.05 | 0.11 | 0.41 |
| CURVED | 1.00 | 0.93 | 1.00 | 0.87 | 0.93 | 1.00 | 0.71 | 0.70 | 0.74 | 0.77 | 0.71 | 0.79 |
| UNKNOWN | 0.84 | 0.85 | 0.97 | 0.88 | 0.85 | 0.90 | 0.62 | 0.59 | 0.66 | 0.63 | 0.64 | 0.66 |
| Average | 0.85 | 0.79 | 0.89 | 0.62 | 0.63 | 0.71 | 0.72 | 0.73 | 0.84 | 0.60 | 0.63 | 0.73 |

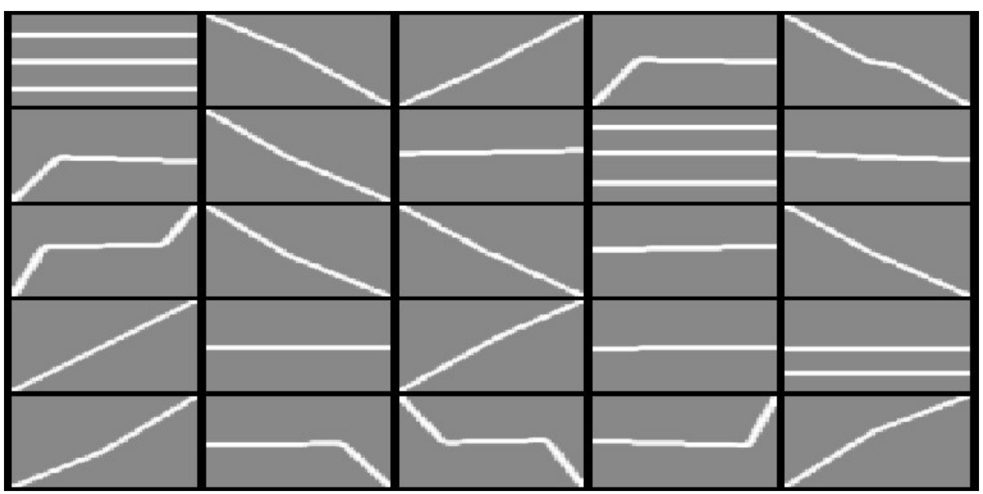
- Motivations and Importance of the problem.
- Introduction and novel contributions.
- Data synthesis in data space.
- Learning from synthetic data.
- Eliminating synthetic gap.
- Data synthesis in feature space.
- Conclusion.

# Eliminate Synthetic Gap

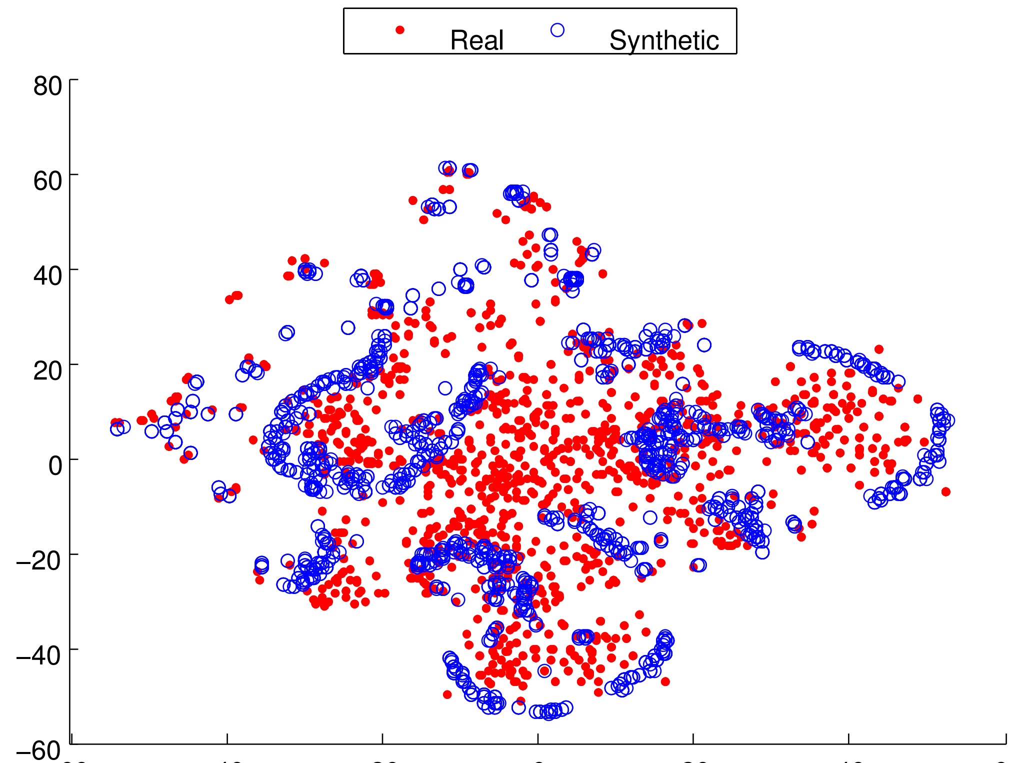
- Synthetic gap.



Actual data



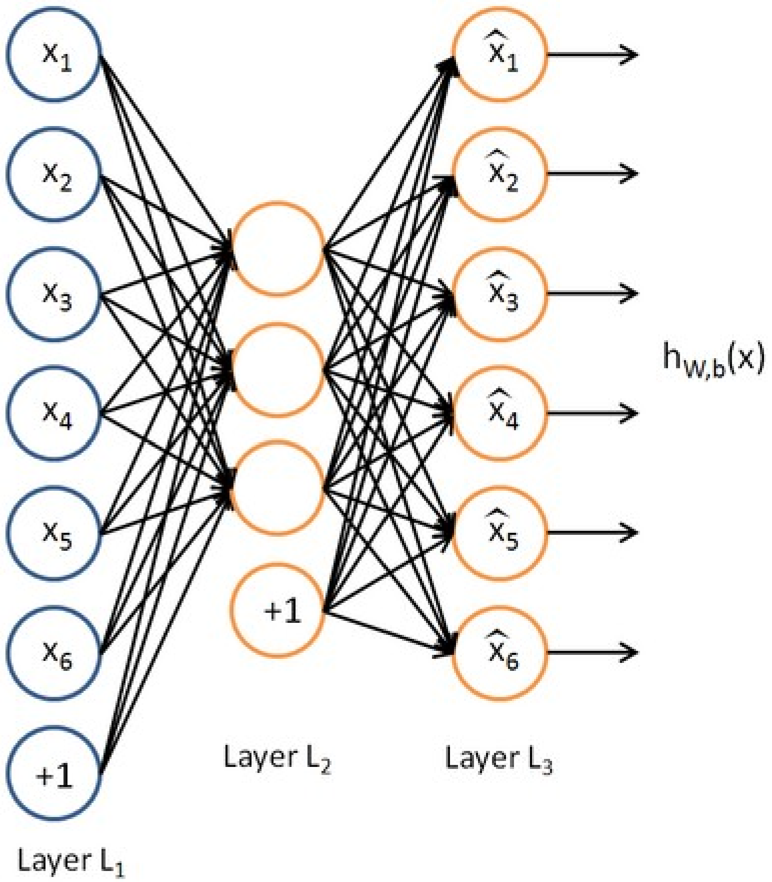
Synthetic data

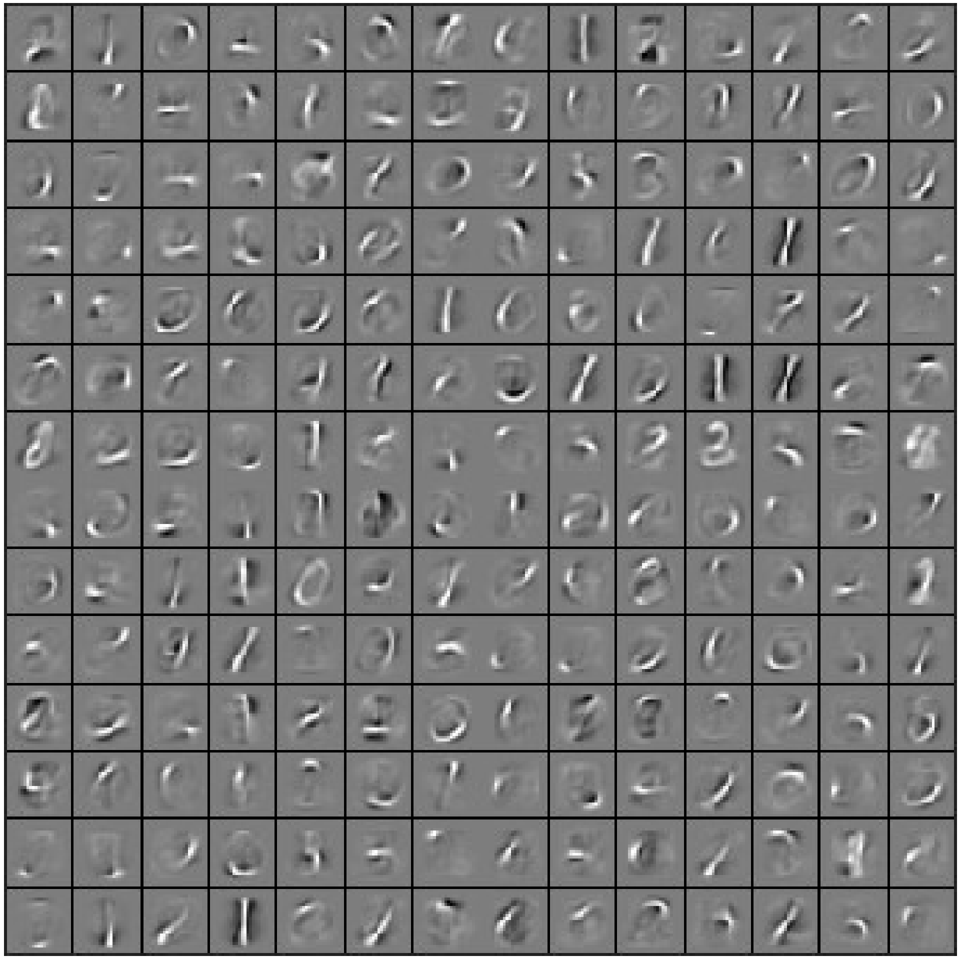


- Previous work.
  - Eliminate distribution means in Kernel Reproducing Hilbert Space (KRHS). [12][46]
  - Subspace alignment. [35][24]
  - Deep neural network using KRHS as domain loss. [32][19]

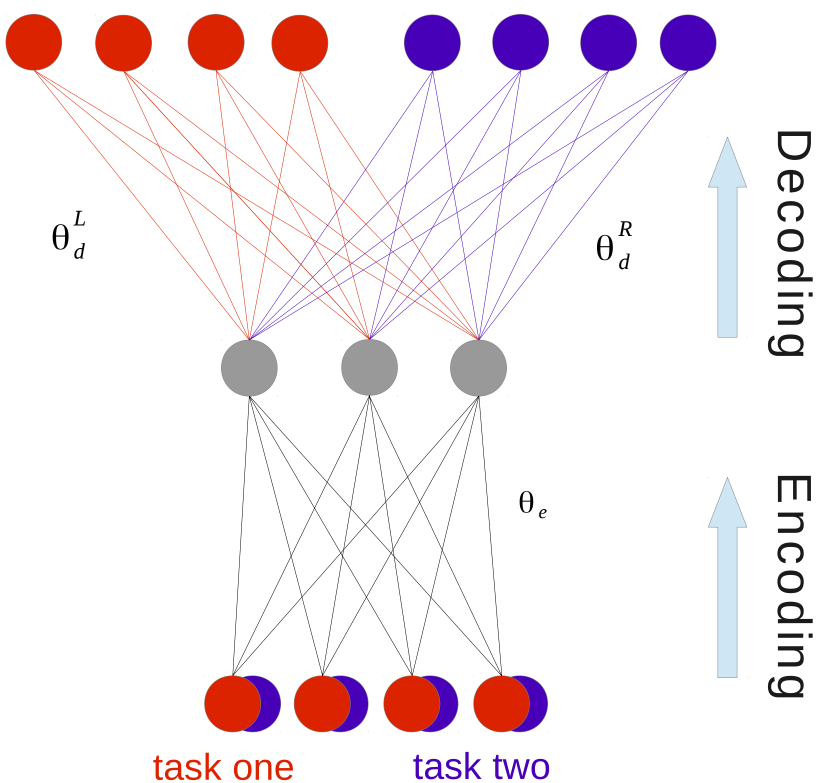
- The proposed approach.
  - Multi-Channel Autoencoder (MCAE)

- Traditional autoencoder

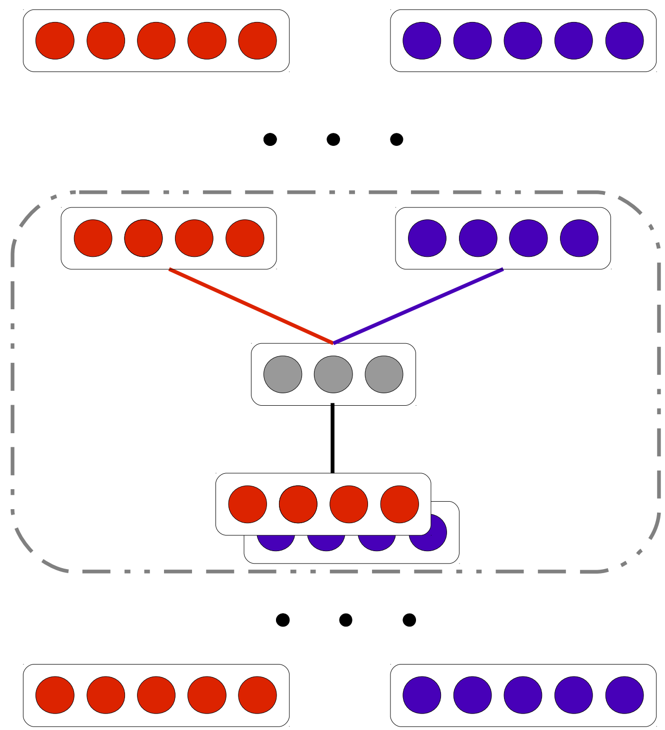


- The proposed MCAE




Basic structure



Stacked up

- MCAE

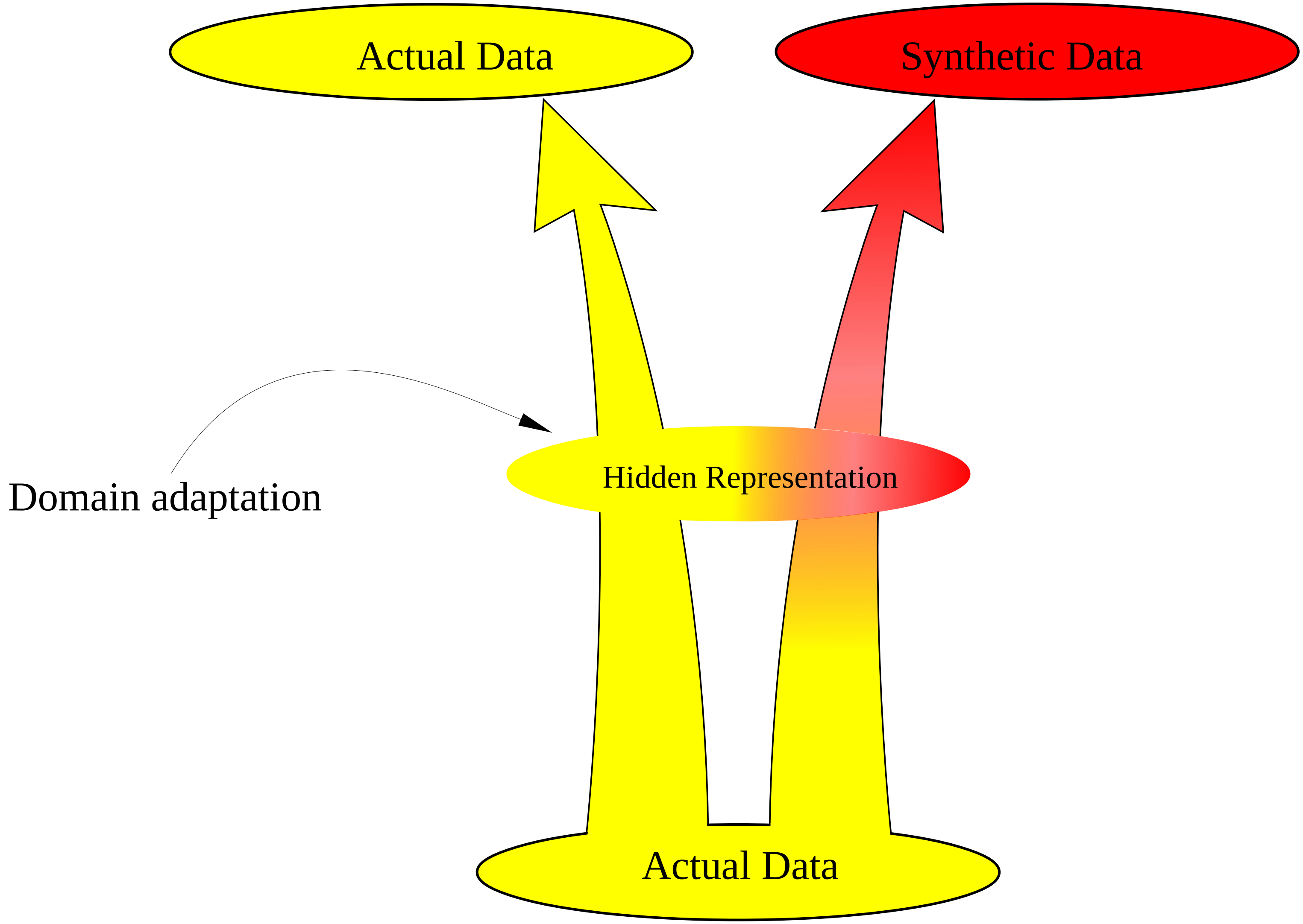
Jointly learn two tasks, left and right, together.

$$E = J^L(\theta_e, \theta_d^L) + J^R(\theta_e, \theta_d^R) + \gamma\Psi$$

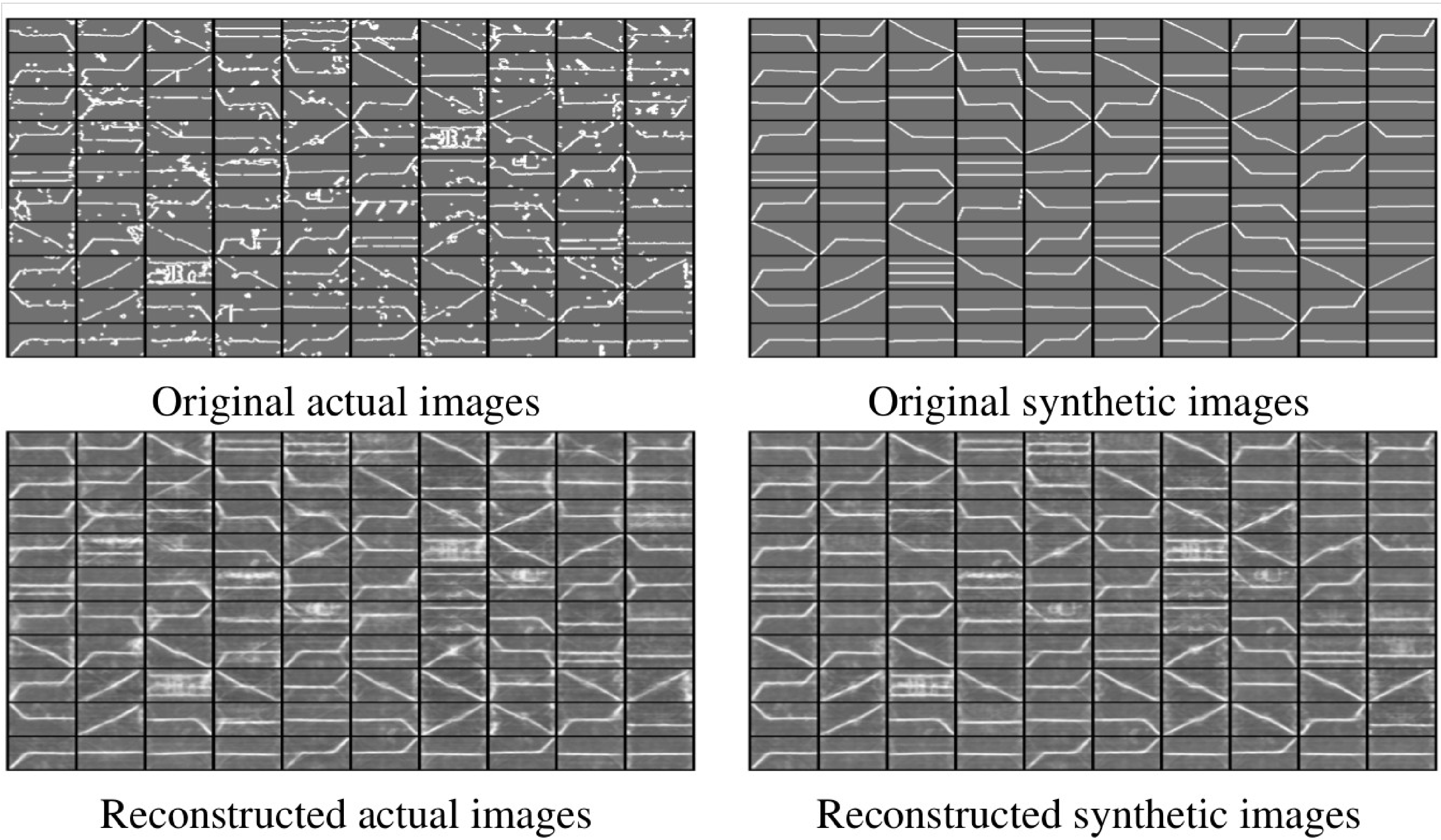
where

$$\Psi = \frac{1}{2}(J^L(\theta_e, \theta_d^L) - J^R(\theta_e, \theta_d^R))^2$$

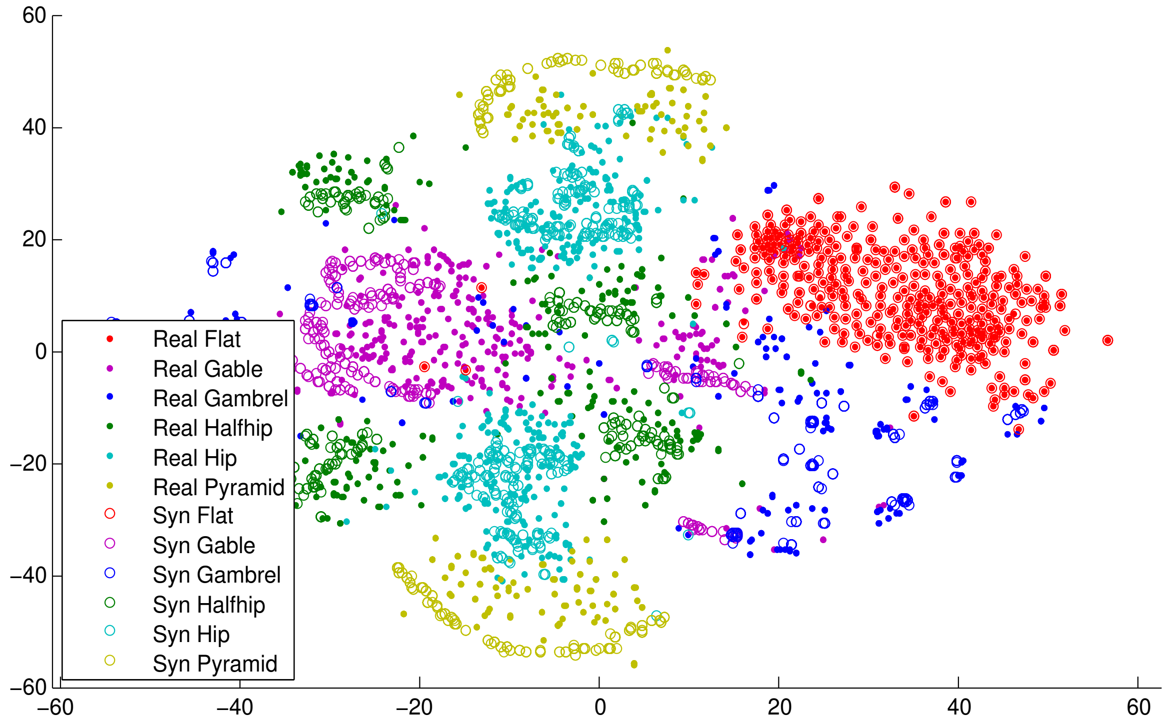
- Configuration of MCAE



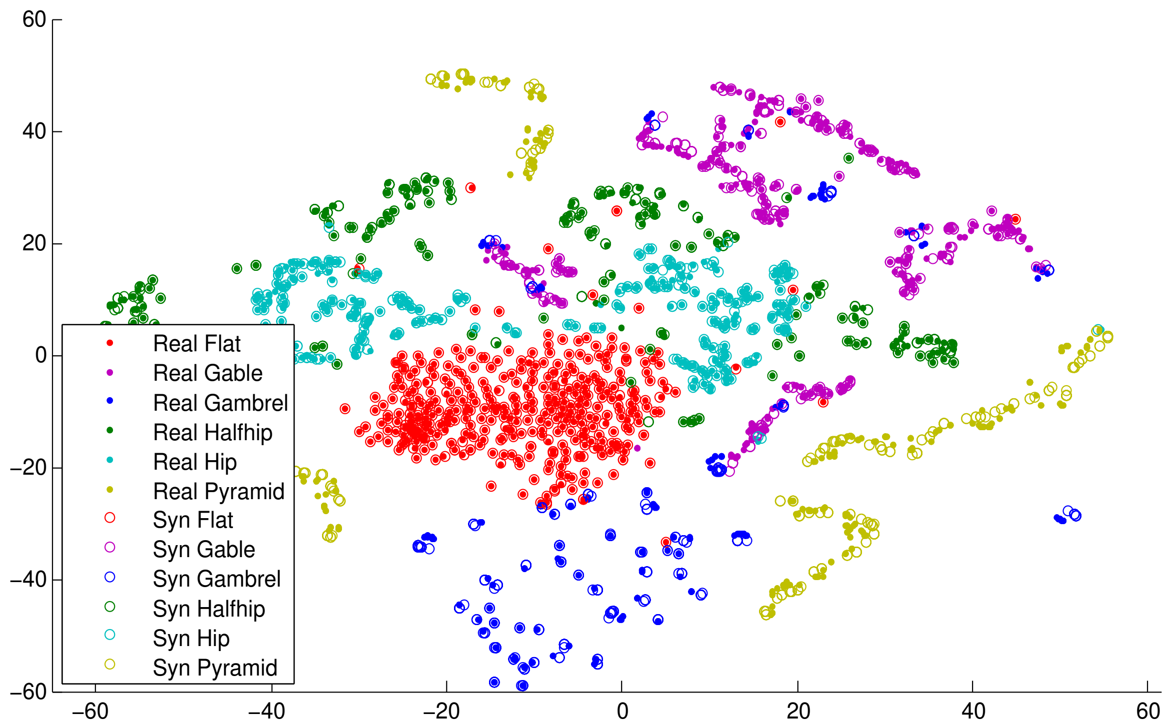
- Visualization of domain adaptation.



Without domain adaptation




With domain adaptation




ILLINOIS INSTITUTE OF TECHNOLOGY

- Classification results

| | Data to train autoencoder | CNN Reconstructed | SVM Encoded |
|---|---|---|---|
| **MCAE** | $\langle$i:*Syn I*, t:Real$\rangle^L$ $\langle$i:Real, t:Real$\rangle^R$ | **0.68** | **0.80** |
| **CIAE** | $\langle$i:*Syn I* + Real, t:*Syn I* + Real$\rangle$ | 0.68 | 0.78 |
| **SAE** | $\langle$i:*Syn I*, t:*Syn I*$\rangle$ | 0.63 | 0.59 |
| **SAE** | $\langle$i:Real, t:Real$\rangle$ | 0.62 | 0.62 |

Roof style dataset

| | Data to train autoencoder | CNN Reconstructed | SVM Encoded |
|---|---|---|---|
| **MCAE** | $\langle$i:*Syn I*, t:Real$\rangle^L$ $\langle$i:Real, t:Real$\rangle^R$ | **0.98** | **0.96** |
| **CIAE** | $\langle$i:*Syn I* + Real, t:*Syn I* + Real$\rangle$ | 0.97 | 0.96 |
| **SAE** | $\langle$i:*Syn I*, t:*Syn I*$\rangle$ | 0.94 | 0.91 |
| **SAE** | $\langle$i:Real, t:Real$\rangle$ | 0.95 | 0.65 |

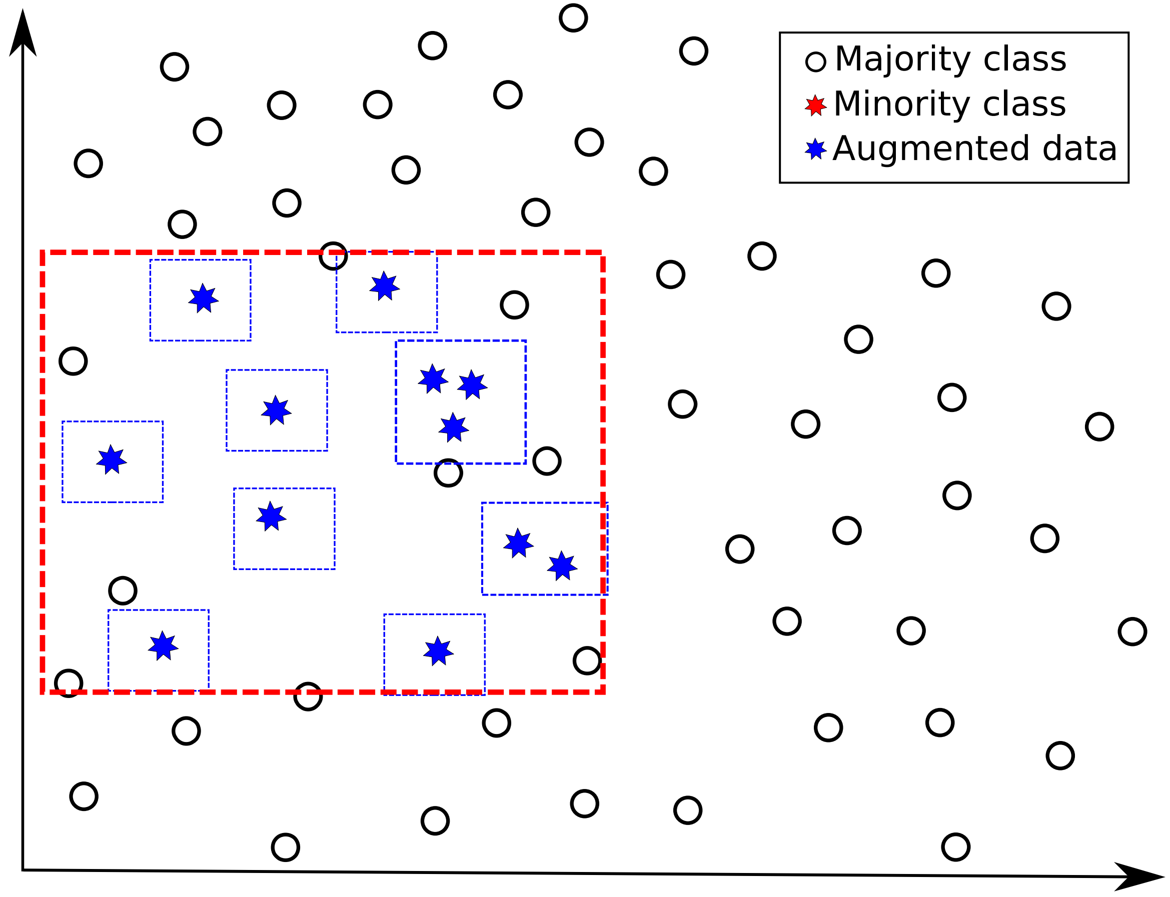
Handwritten digit dataset

- Motivations and Importance of the problem.
- Introduction and novel contributions.
- Data synthesis in data space.
- Learning from synthetic data.
- Eliminating synthetic gap.
- Data synthesis in feature space.
- Conclusion.

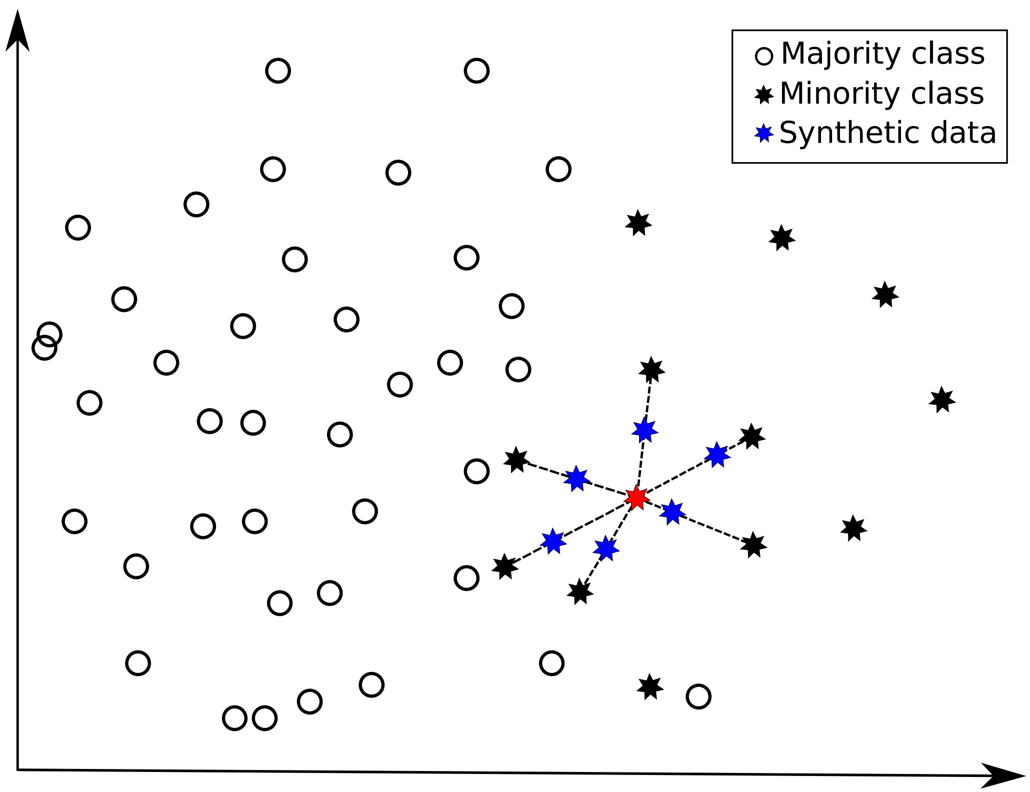
# Data Synthesis in Feature Space

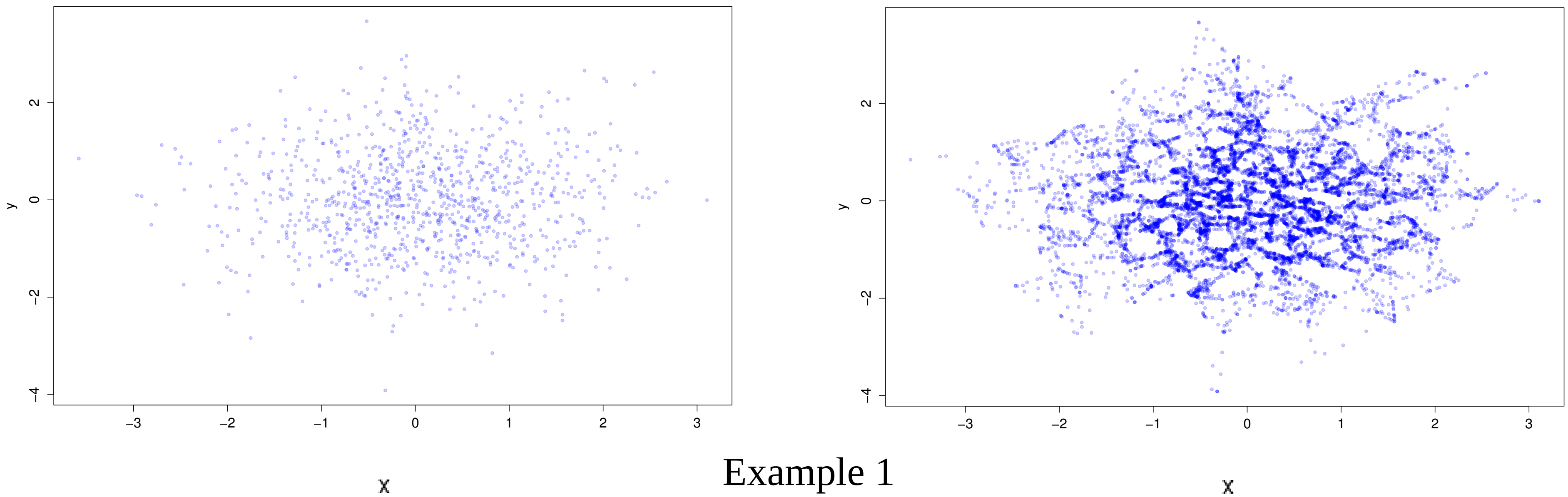
- Why in feature space:
  - Limitation for data synthesis in data space.
  - An universal approach for most problems.
  - Sometimes easier to detect patterns and relationship between data.
- Challenges:
  - Data representation is too abstract in feature space. To derive more data has to look at entire dataset.

- Previous work.
  - Creating identical samples.

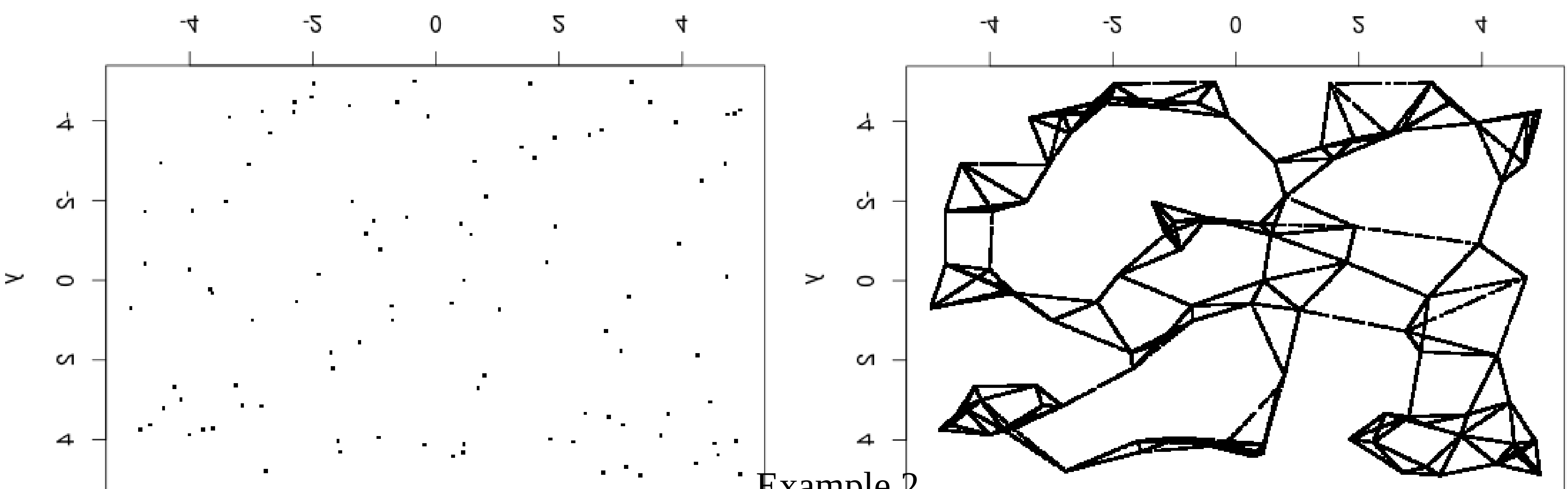


- Previous work.
  - Creating identical samples.
  - Synthetic minority oversampling technique (SMOTE)






Example 1

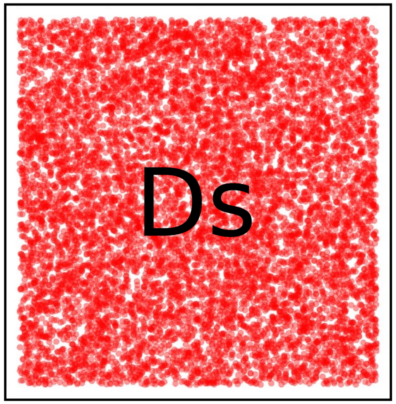


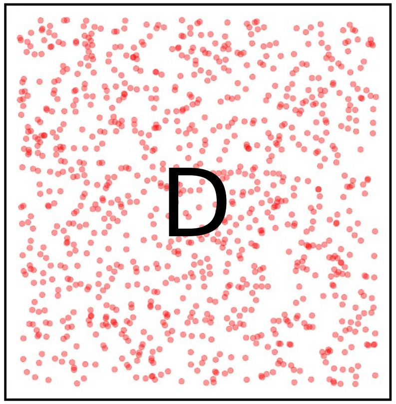
Example 2

- The proposed approach.
    - Up-sampling by keeping global structure of the dataset.

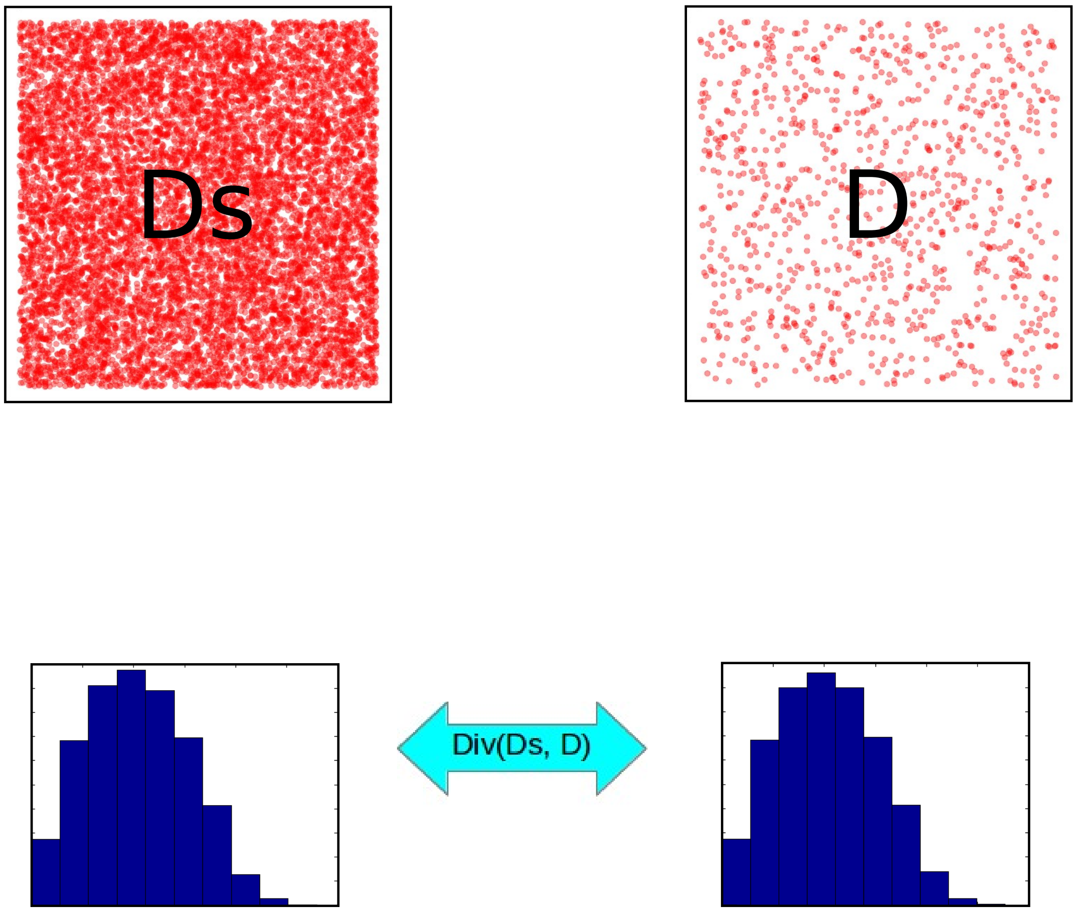
      Suppose we are given a dataset called $D$, the goal is to find out a dataset $Ds$ which is a up-sampled version of $D$

- The proposed approach.

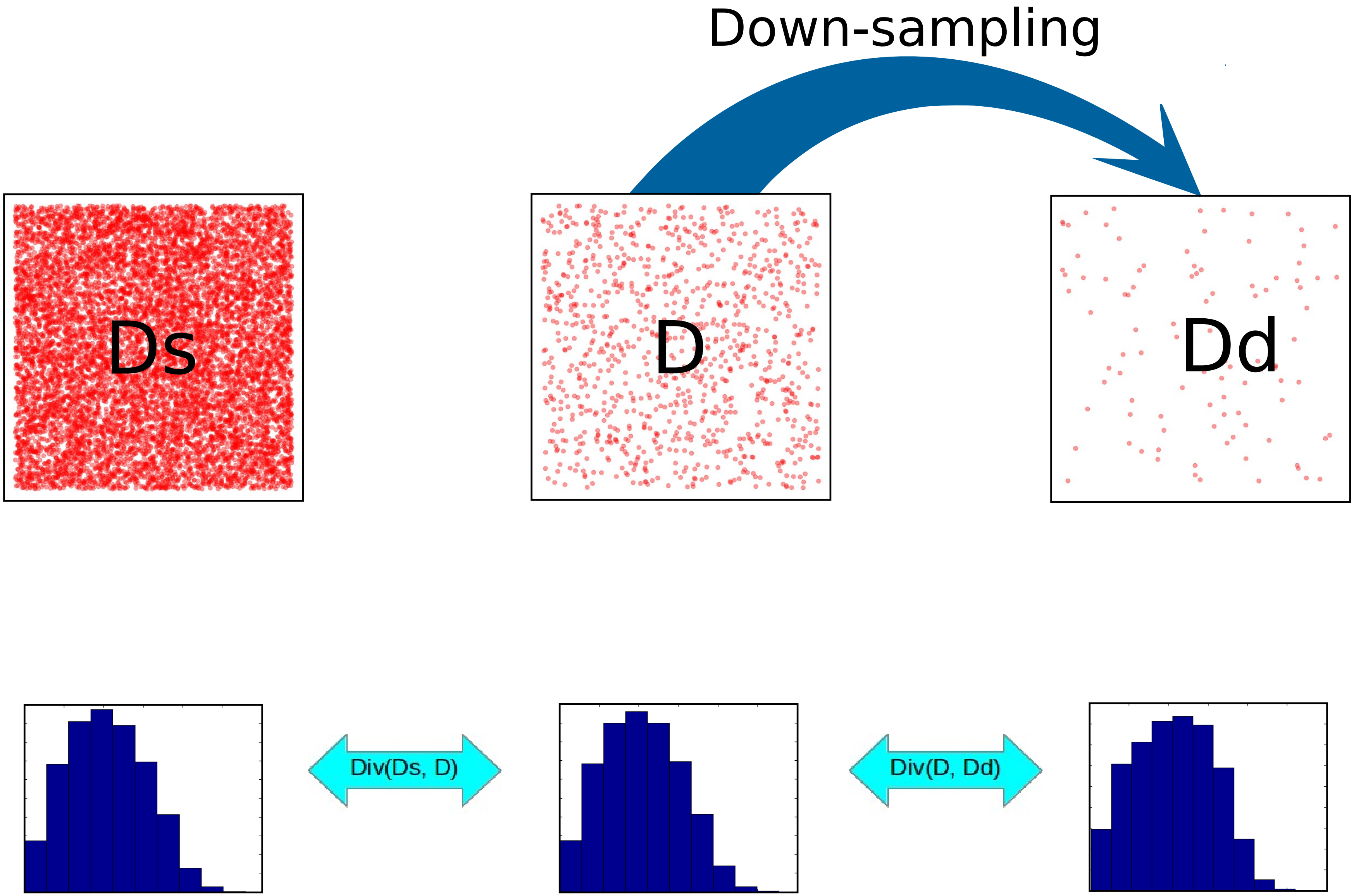


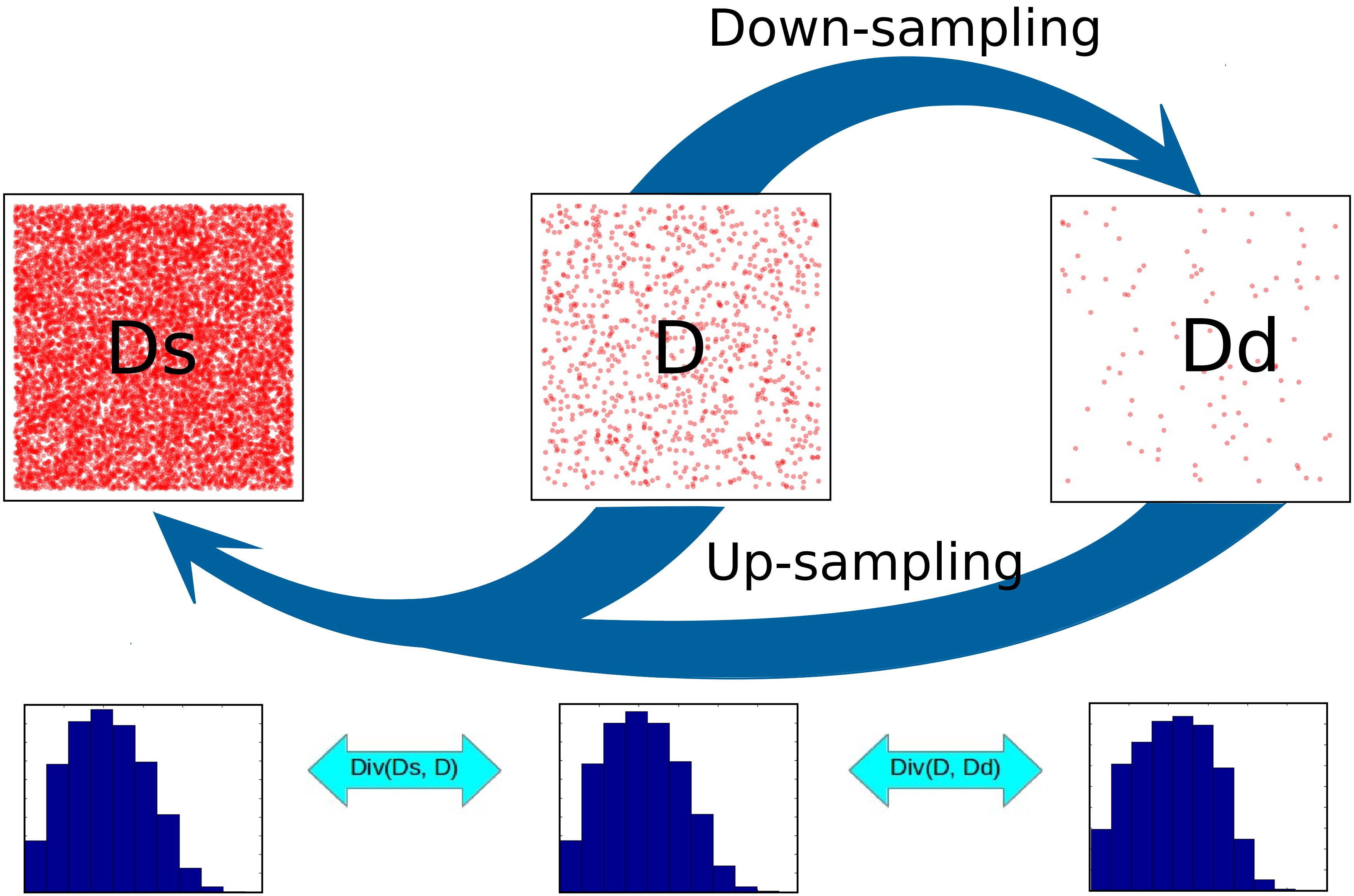
- The proposed approach.



- The proposed approach.



- The proposed approach.



- The proposed approach.

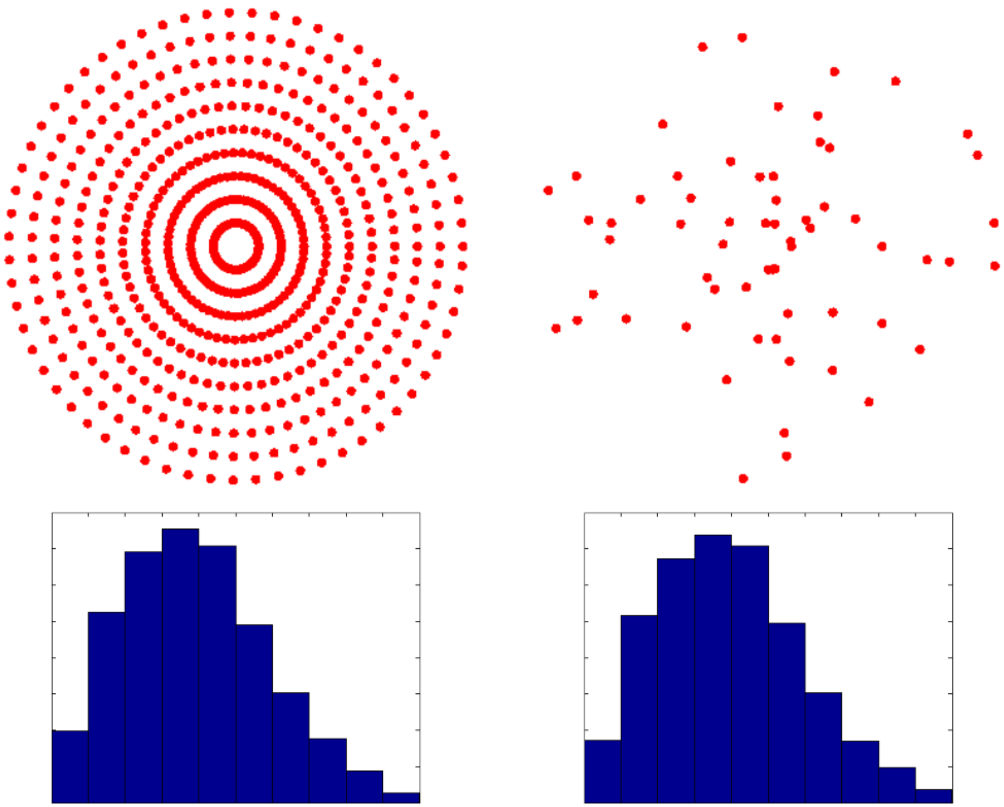
  1) Down-sampling: Minimize Kullback-Leibler divergence between two datasets.

$$E = Div(H(D), H(D_d)) = \sum_{i}^{k} b_{\mathrm{D}}^{i} \ln \frac{b_{\mathrm{D}}^{i}}{b_{\mathrm{Dd}}^{i}} + b_{\mathrm{Dd}}^{i} \ln \frac{b_{\mathrm{Dd}}^{i}}{b_{\mathrm{D}}^{i}}$$

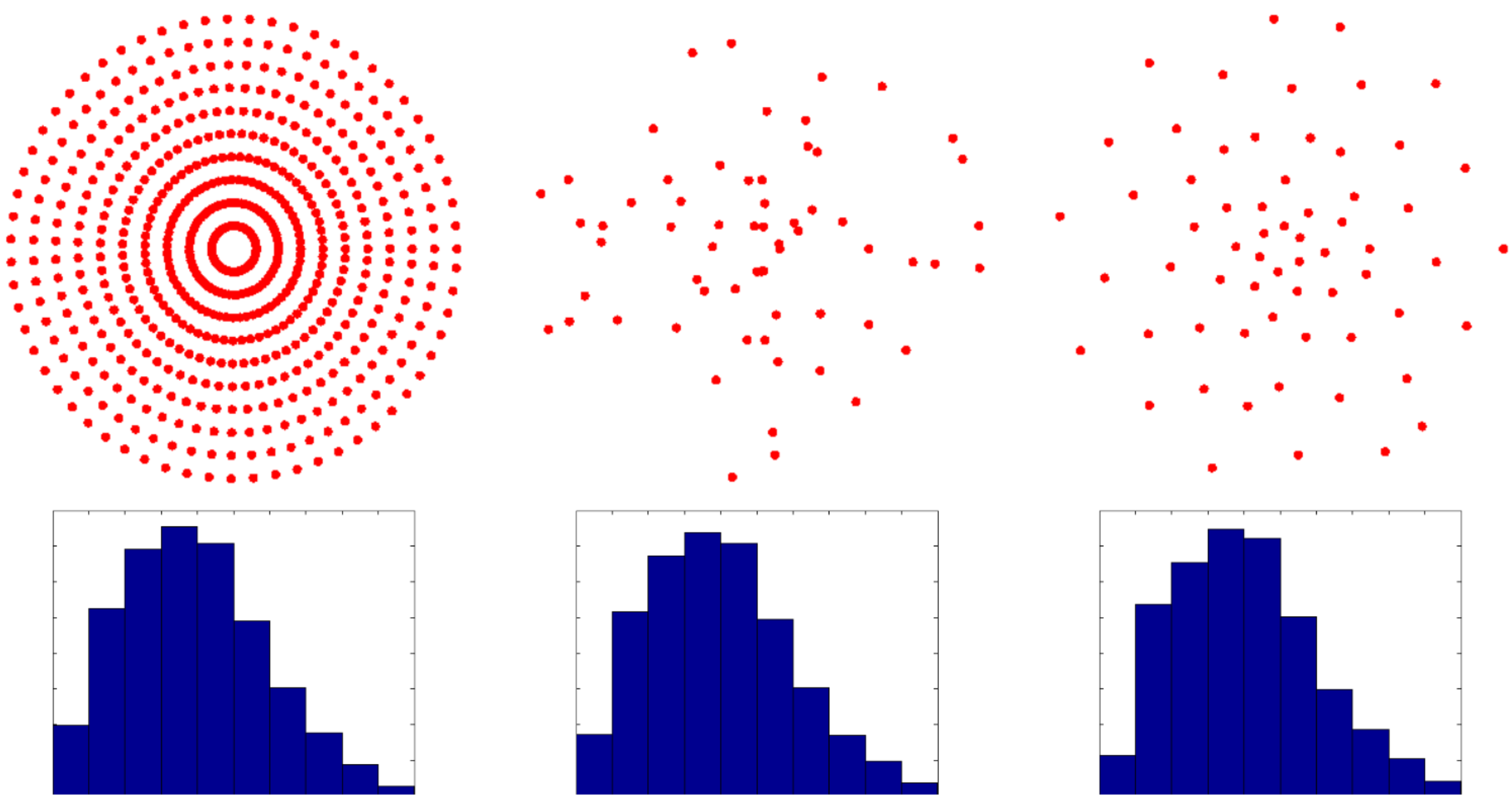
- Down-sampling

Minimize Kullback-Leibler divergence between two datasets.

$$E = Div(H(D), H(D_d)) = \sum_{i}^{k} b^i_{\mathrm{D}} \ln \frac{b^i_{\mathrm{D}}}{b^i_{\mathrm{Dd}}} + b^i_{\mathrm{Dd}} \ln \frac{b^i_{\mathrm{Dd}}}{b^i_{\mathrm{D}}}$$



$$E = Div(H(D), H(D_d)) = \beta \sum_{i}^{k} b^i_{\mathrm{D}} \ln \frac{b^i_{\mathrm{D}}}{b^i_{\mathrm{Dd}}} + b^i_{\mathrm{Dd}} \ln \frac{b^i_{\mathrm{Dd}}}{b^i_{\mathrm{D}}} + \underbrace{(1-\beta) \sum_{i}^{n} \left( \frac{k^i_{\mathrm{Dd}}}{|N_{d^i_{\mathrm{D}}}|} - r^{\downarrow}_{\mathrm{DDd}} \right)^2}_{\text{Sampling adaptively}} \quad (5.2)$$



• Comparison of classification results.

| KNN (k=5) | fps | kmeans | kmeansrand | rand | my |
|---|---|---|---|---|---|
| Abalone | 0.0908 | 0.1366 | 0.1427 | 0.1405 | 0.1961 |
| Banknote | 0.9963 | 0.9963 | 1 | 0.9993 | 1 |
| Blood | 0.4973 | 0.6049 | 0.6001 | 0.5929 | 0.6171 |
| Bcancer | 0.6139 | 0.6355 | 0.6444 | 0.638 | 0.6424 |
| CT | 0.9731 | 0.9683 | 0.9715 | 0.9661 | 0.967 |
| CMS | 0.4778 | 0.4778 | 0.4773 | 0.4778 | 0.4778 |
| CNAE-9 | 0.7051 | 0.7445 | 0.7699 | 0.819 | 0.8367 |
| Diabetes | 0.6215 | 0.7053 | 0.7053 | 0.7082 | 0.7254 |
| Haberman | 0.4841 | 0.6325 | 0.622 | 0.6331 | 0.6431 |
| Iono | 0.4076 | 0.5483 | 0.8036 | 0.8387 | 0.8105 |
| V.S. | 0.6643 | 0.6585 | 0.65 | 0.6433 | 0.6704 |
| Wine | 0.9155 | 0.9193 | 0.9221 | 0.9161 | 0.9196 |

| RANDFOREST | fps | kmeans | kmeansrand | rand | my |
|---|---|---|---|---|---|
| Abalone | 0.5658 | 0.5801 | 0.5899 | 0.5379 | 0.5838 |
| Banknote | 0.9999 | 0.9994 | 0.991 | 0.9917 | 0.9971 |
| Blood | 0.7024 | 0.7069 | 0.7337 | 0.7301 | 0.7573 |
| Bcancer | 0.6482 | 0.6528 | 0.6523 | 0.6493 | 0.6562 |
| CT | 0.9948 | 0.9907 | 0.9905 | 0.9863 | 0.9873 |
| CMS | 0.7433 | 0.7439 | 0.7674 | 0.7482 | 0.7603 |
| CNAE-9 | 0.946 | 0.9549 | 0.925 | 0.9291 | 0.9637 |
| Diabetes | 0.7975 | 0.8413 | 0.827 | 0.8202 | 0.8359 |
| Haberman | 0.7052 | 0.7749 | 0.7143 | 0.7295 | 0.7627 |
| Iono | 0.8706 | 0.9396 | 0.9535 | 0.9502 | 0.9562 |
| V.S. | 0.825 | 0.8445 | 0.8299 | 0.8278 | 0.8468 |
| Wine | 0.9972 | 0.9947 | 0.994 | 0.994 | 0.9952 |

- Down-sampling

Minimize Kullback-Leibler divergence between two datasets.

$$E = Div(H(D), H(D_d)) = \beta \sum_{i}^{k} b_{\mathrm{D}}^{i} \ln \frac{b_{\mathrm{D}}^{i}}{b_{\mathrm{Dd}}^{i}} + b_{\mathrm{Dd}}^{i} \ln \frac{b_{\mathrm{Dd}}^{i}}{b_{\mathrm{D}}^{i}} + (1-\beta) \sum_{i}^{n} \left( \frac{k_{\mathrm{Dd}}^{i}}{|N_{d_{\mathrm{D}}^{i}}|} - r_{\mathrm{DDd}}^{\downarrow} \right)^2 \quad (5.2)$$

- Up-sampling:

$$H(D_s) = H(D) + \alpha(H(D) - H(D_d))$$

# Conclusion