

Data Synthesis For Object Recognition

Xi Zhang

Advisor: Professor Agam

xzhang22@hawk.iit.edu

Computer Science Department

Illinois Institute of Technology

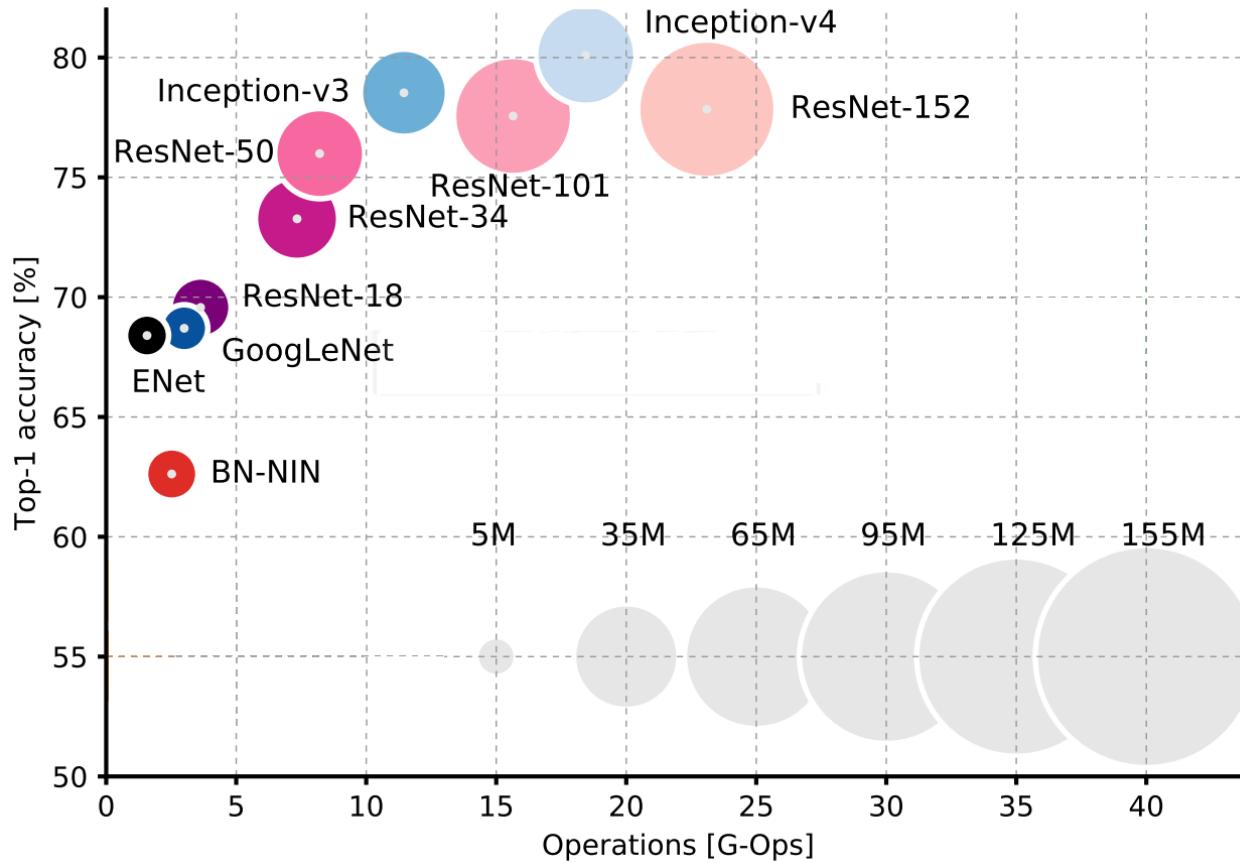
Overview

- ❖ Motivation.
- ❖ 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.

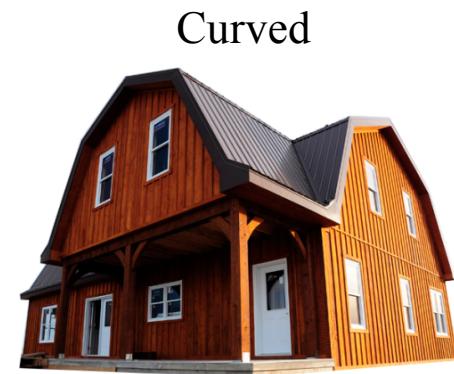
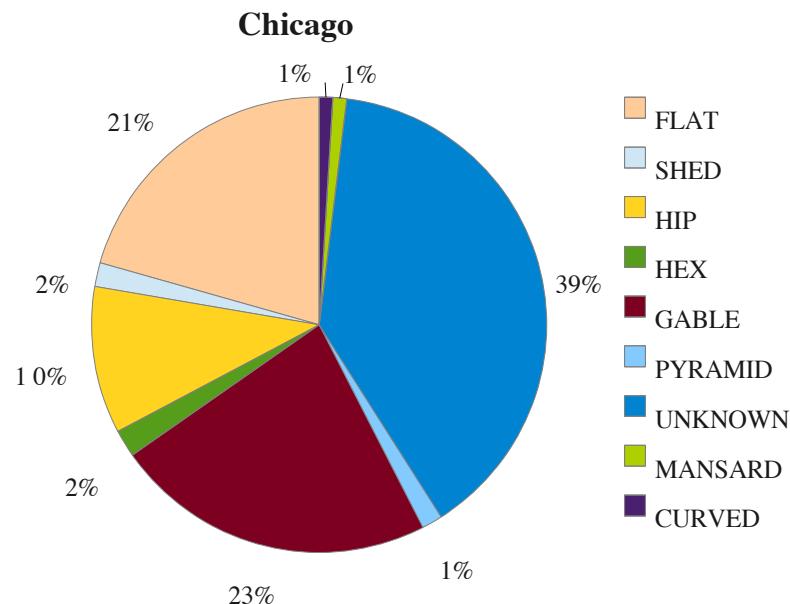


Motivation



Model complexity V.S. Accuracy On ImageNet dataset

- 1) To gain a good performance of a machine learning process, more high quality data is always desired.
- ❖ Rare cases (absolute rarity)
 - ❖ Rare classes (relative rarity)



2) Supervised learning requires high quality labeled data.

- Time consuming, expensive.
- Sometimes, impossible

2) Supervised learning requires high quality labeled data.

- Time consuming, expensive.
- Sometimes, impossible



2D optical flow



3D point cloud

- ❖ 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.



Solution:

Synthesizing data which will be used as training data.

Challenges:

- Where to synthesize? (As image or as features)
- How to synthesize?
- How to use synthesized data?

Novel contributions:

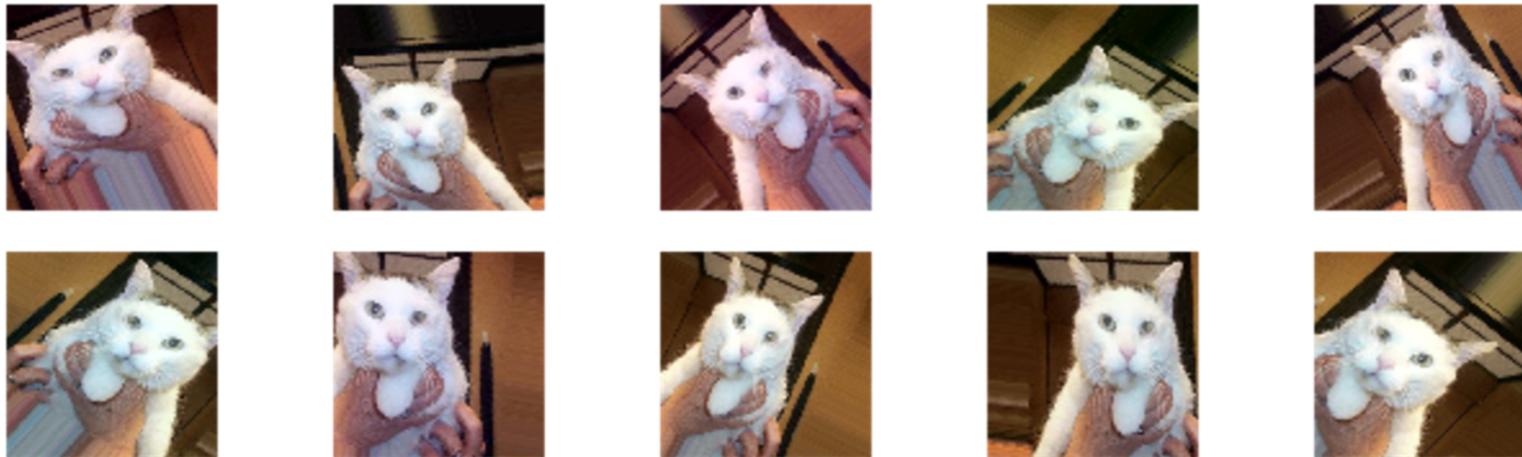
1. Data synthesis in data space.
2. Learning from synthetic data.
3. Eliminating synthetic gap.
4. Data synthesis in feature space.



- ❖ 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.

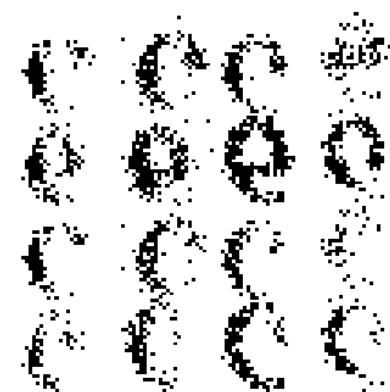
Existing methods.

- ❖ Geometric transformation. [138][139]

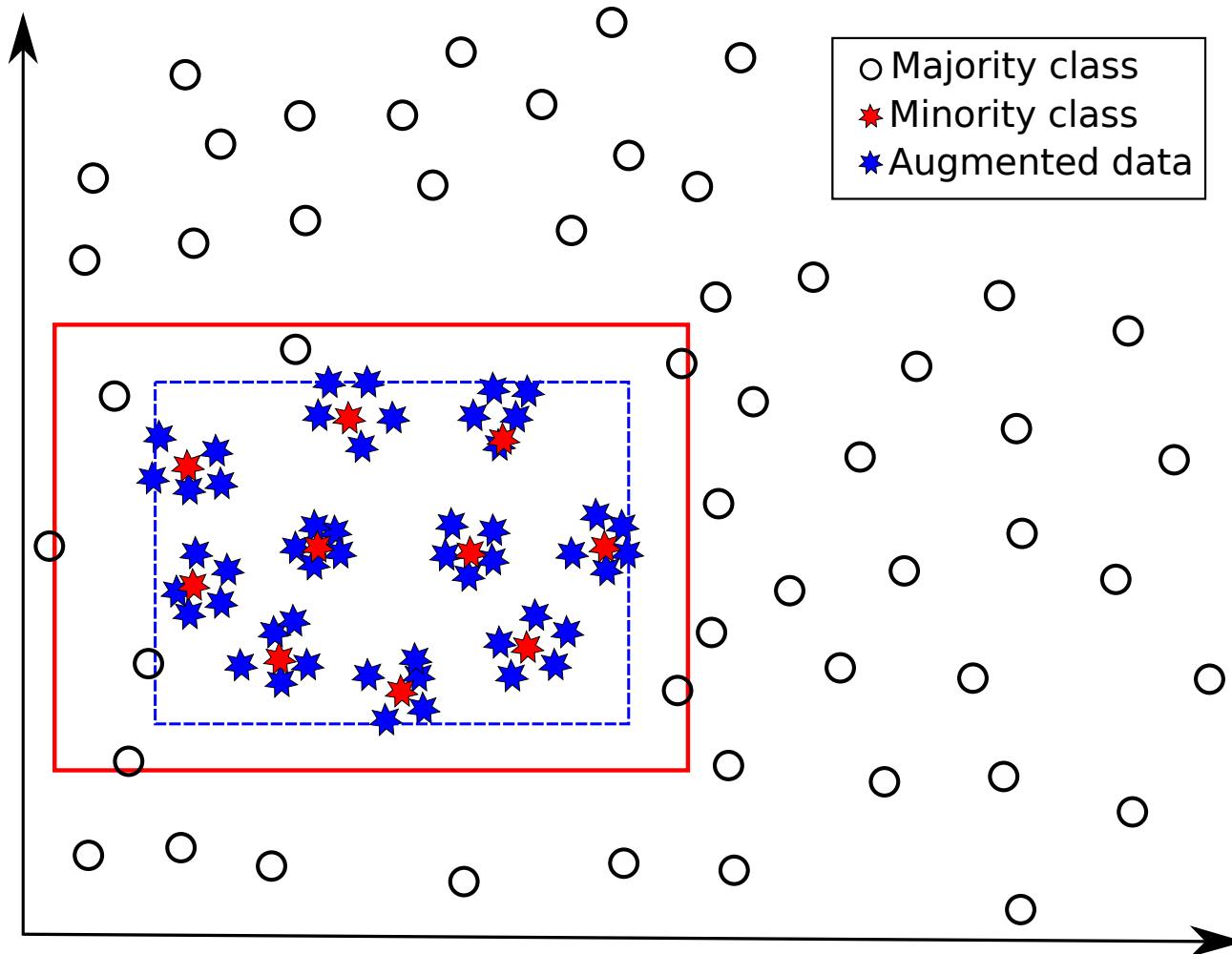


- ❖ Color space transformation. [6][113]

e



- ❖ Disadvantage of existing methods.

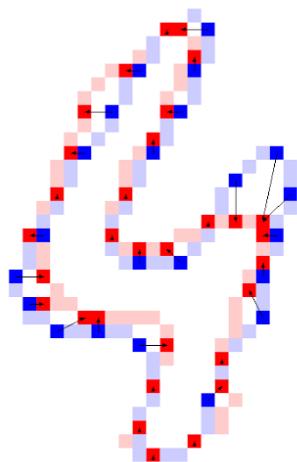


The Proposed Approach

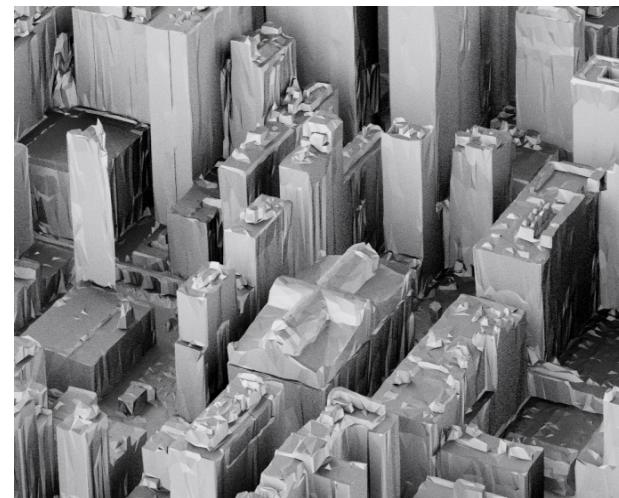
In contrast to existing methods, two approaches are used in my research:

1. Generate a conclusive templates from data.
2. Deep neural networks, which synthesizes data by inferring from unsupervised learning.

Examples of Using Template



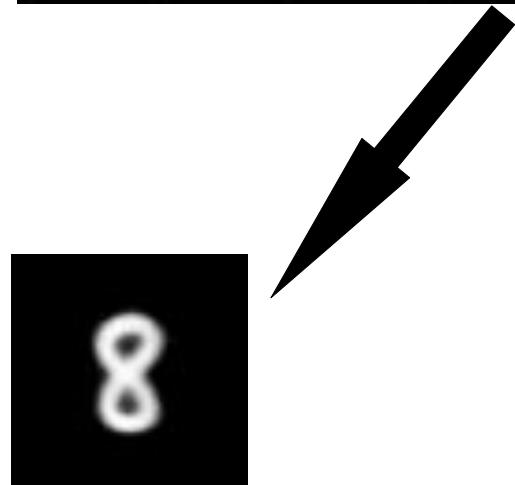
Handwritten digits recognition [25]



Aerial LiDAR roof style classification [20]

Showcase One – Digit Recognition

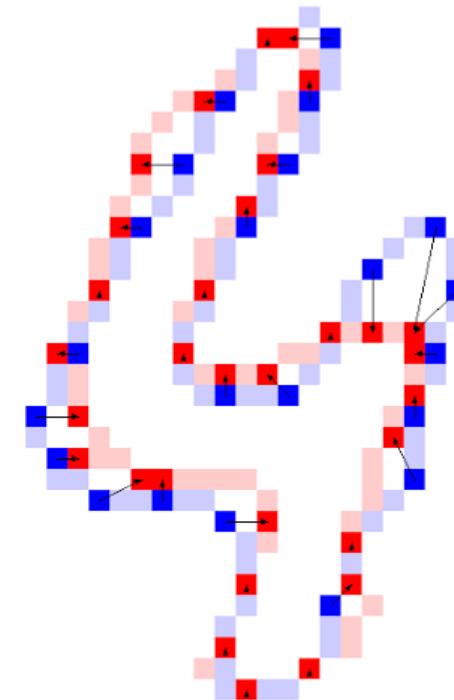
- ❖ Generate prototype by unifying images.



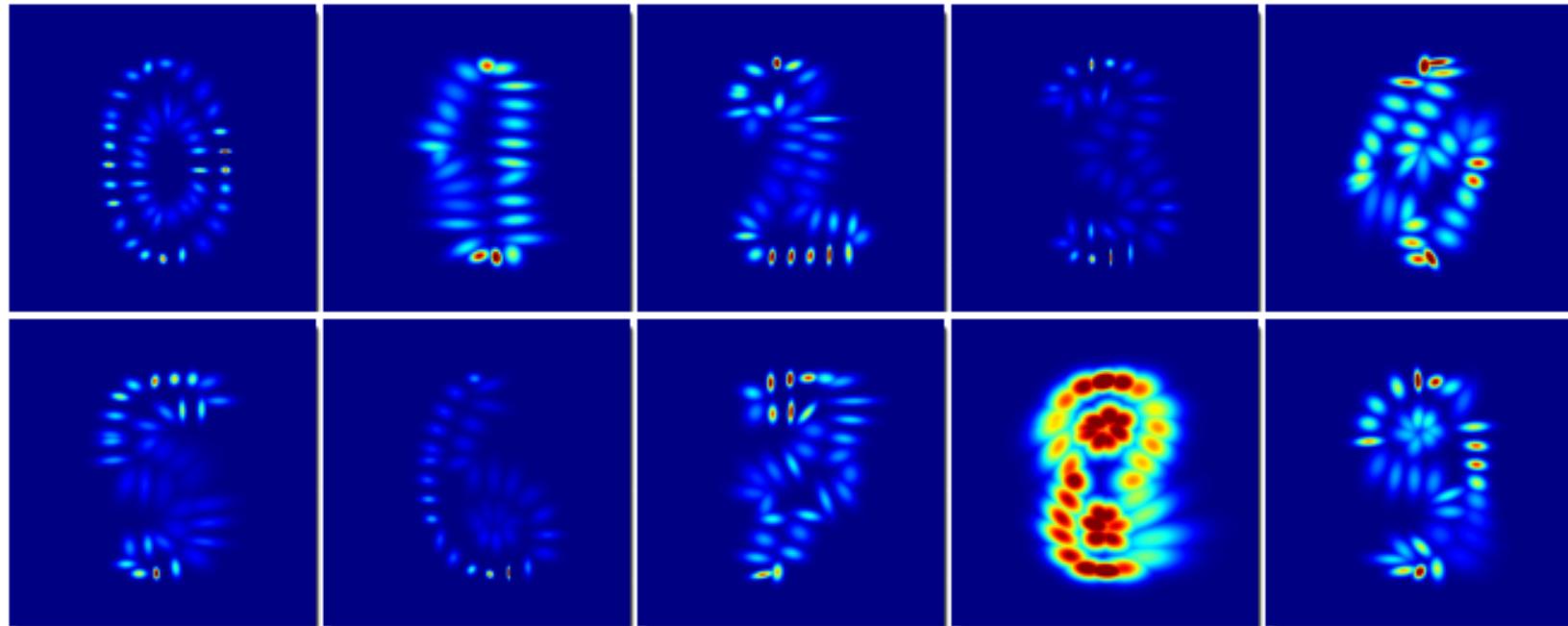
For N images $\{I_i\}_{i=1}^N$ with the same digit, solve for transformations $\{T_i\}_{i=1}^N$, that $\{T_i \cdot I_i\}_{i=1}^N$ minimize joint entropy

Building correspondence among data.

- ❖ Set control points on prototype.
- ❖ Find corresponding control points on each data.



- ❖ Data synthesis by drawing samples from a computed distribution.



- ❖ Data synthesis by interpolation/extrapolation between nearest neighbors.

- ❖ Boost the performance by adding synthetic data.

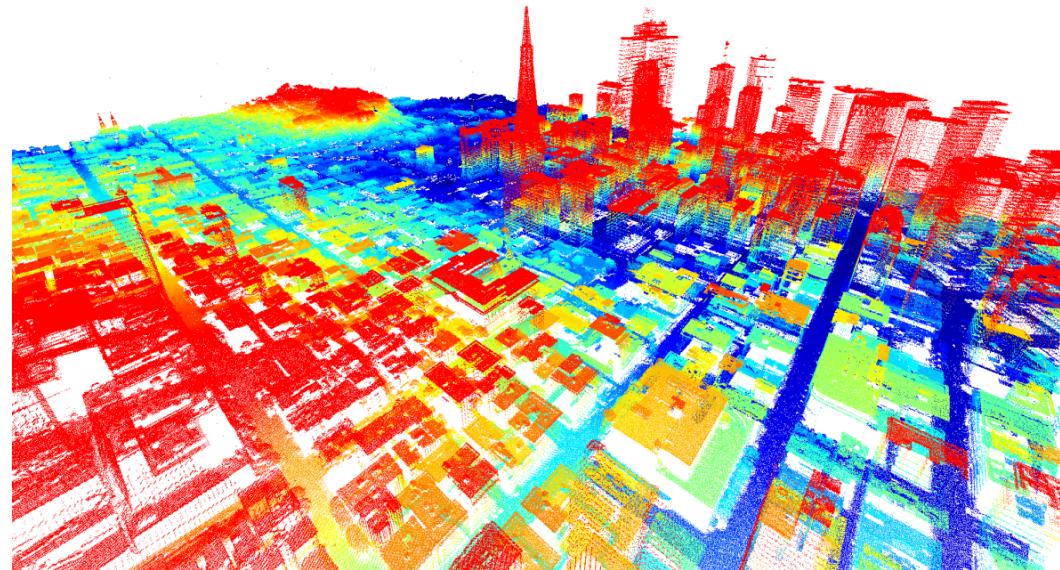
	Real	Syn	Real+Syn
CNN	0.65	0.68	0.70
SVM	0.77	0.78	0.80

F1 score of classification results

Showcase Two – Roof Recognition

Aerial LiDAR

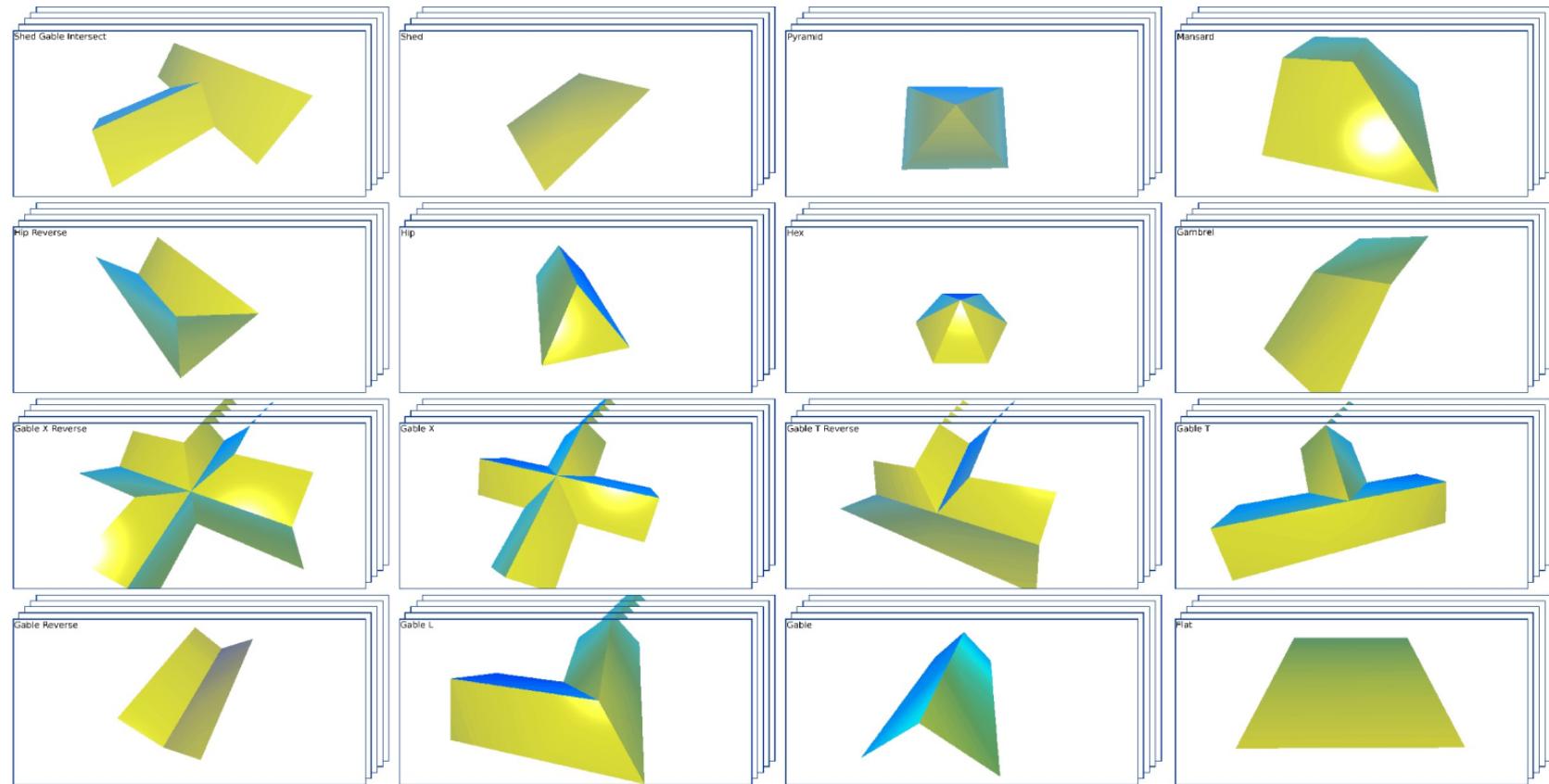
Downtown San Francisco



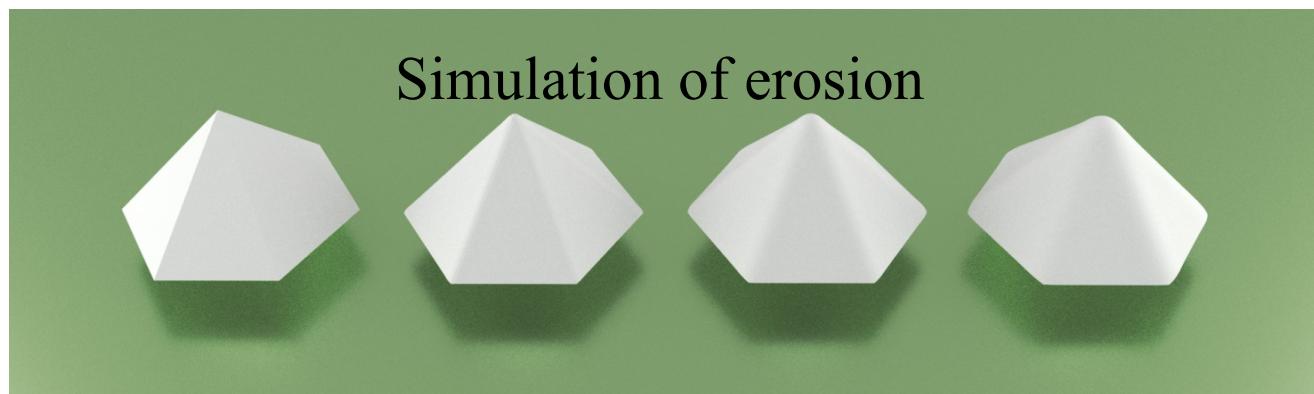
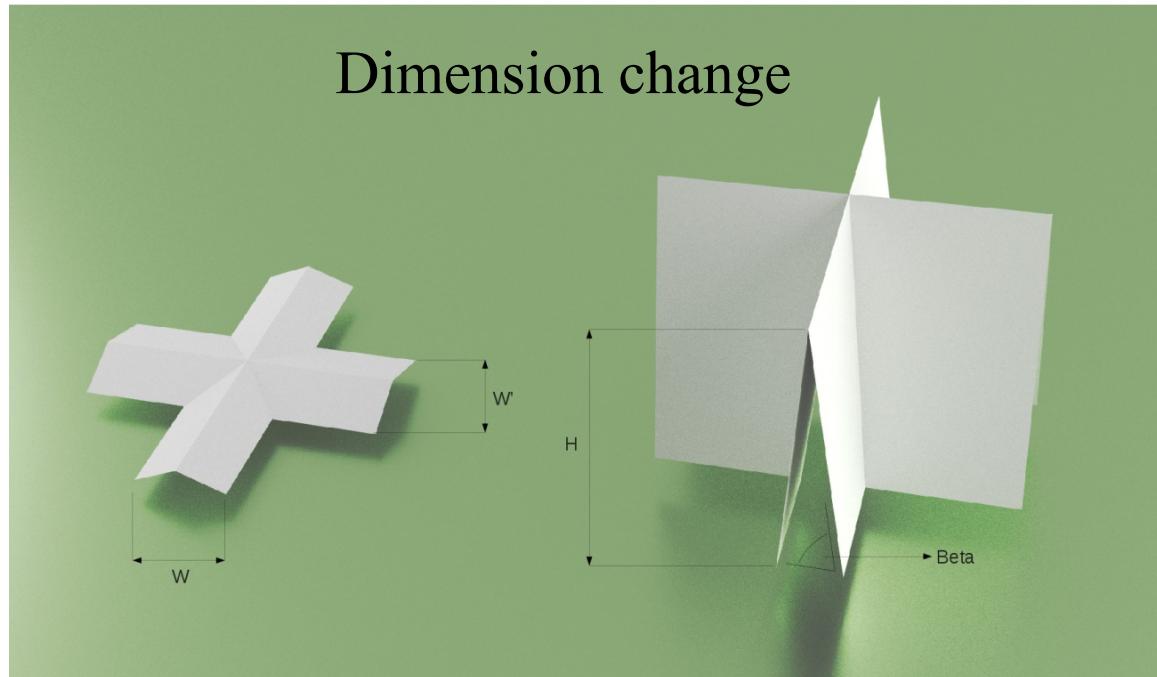
Candidate roof styles



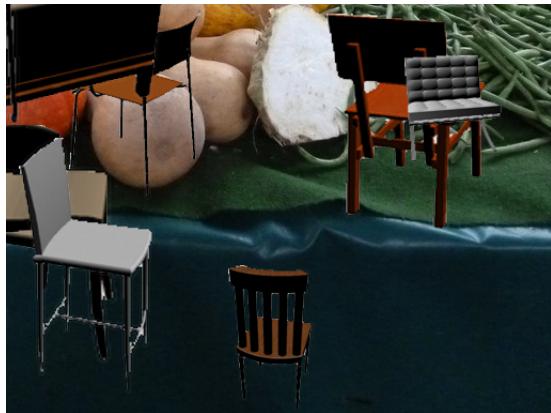
- ❖ Build prototypes.



- ❖ Derive more data.

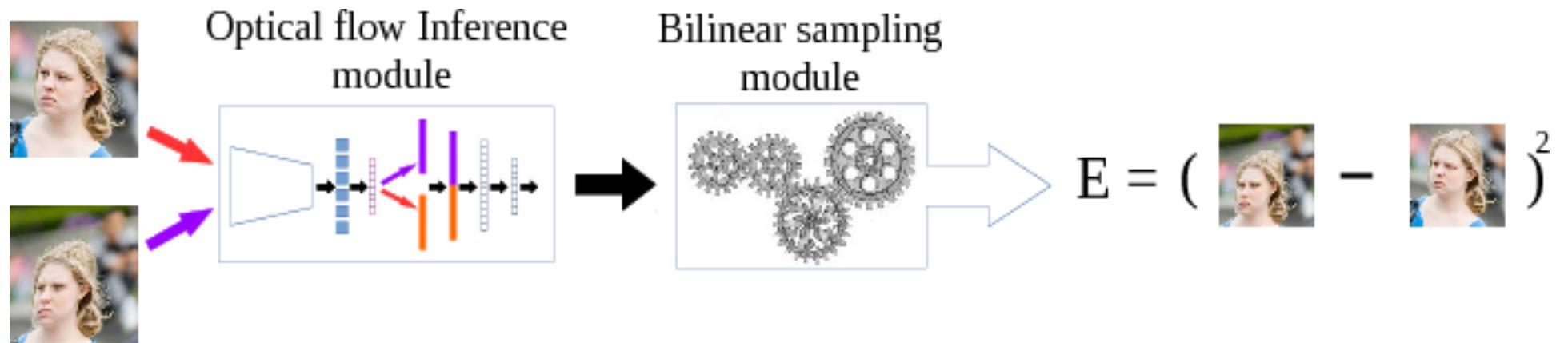


Example of Inferring data from Unsupervised learning

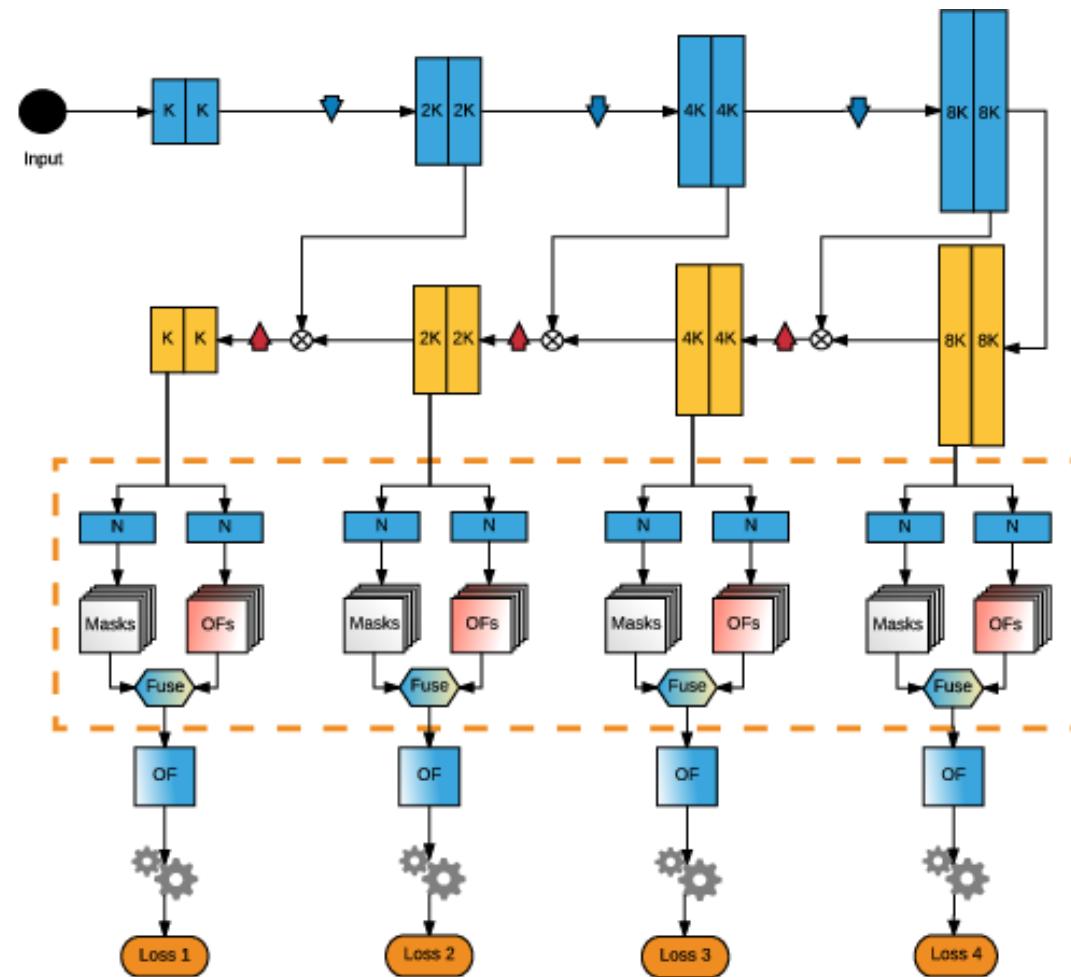


Unsupervised learning of optical flow using neural networks. [23]

- ❖ Framework of unsupervised optical flow learning



❖ Introduction of using mask modules



Conv residual block



Downsampling



Deconv residual block



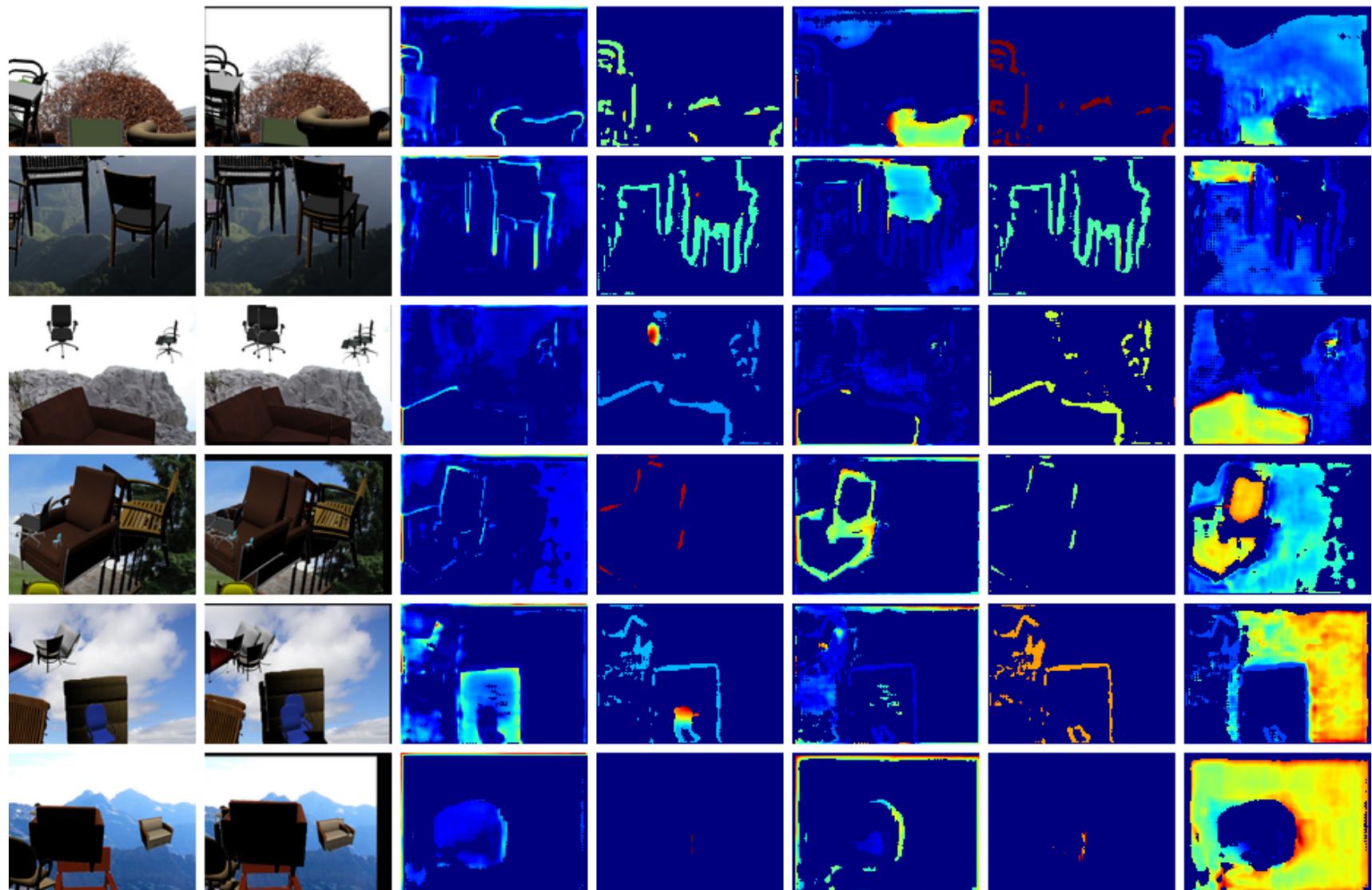
Upsampling



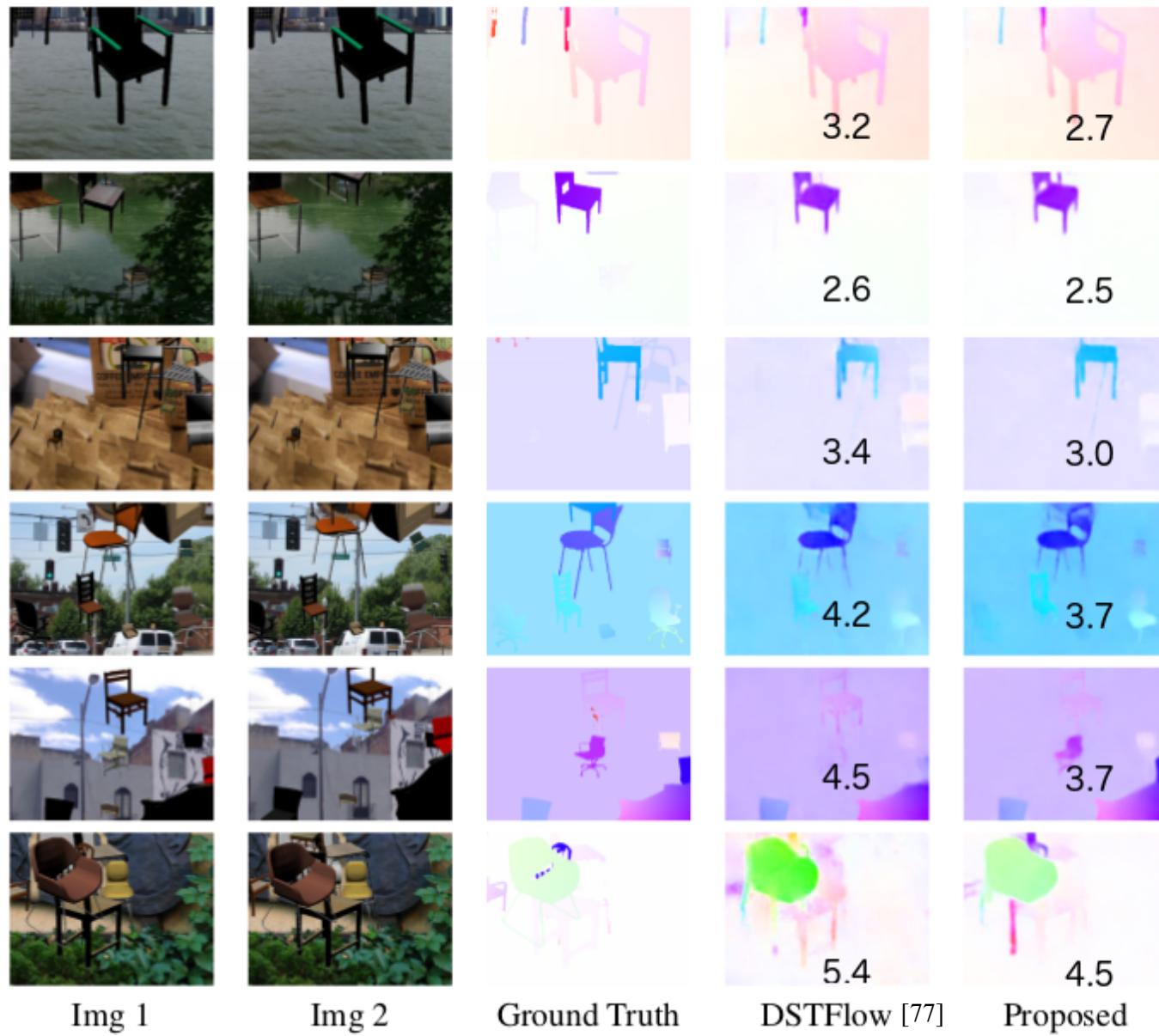
Concatenate



- ❖ Learned masks.



❖ Results



❖ Results

	Chairs	KITTI	SC occ	SC noc	SF occ	SF noc	Time (sec)
Epicflow [†] [87]	2.94	1.52	4.12	1.36	6.29	3.06	16
Deepflow [†] [88]	3.53	1.53	5.38	1.77	7.21	3.34	17
FlowNetS (C+S)*[2]	3.04	5.23	6.96	N.A.	7.76	N.A.	0.08
DSTFlow [77]	5.11	4.02	10.40	5.20	11.11	5.92	0.1
NoMask	5.03	4.11	10.78	6.22	12.32	6.03	0.1
Proposed	4.53	3.67	9.56	4.89	10.21	4.84	0.17

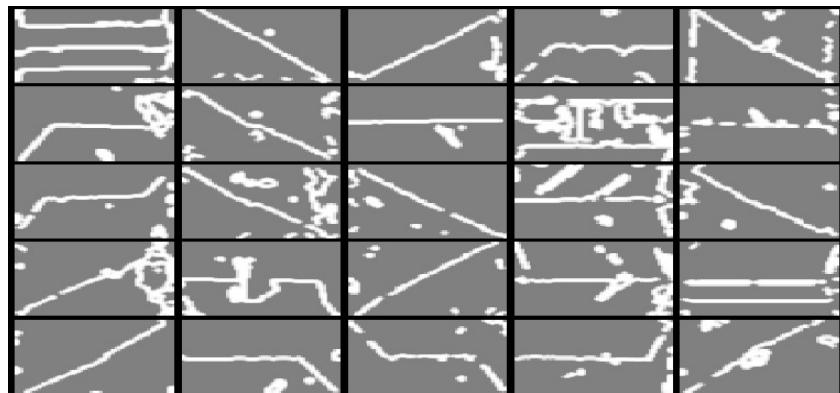
❖ Results

	Train	Fine Tune	Test	EPE
Exp. 1 (sup)	KITTI w. labels	N.A.	KITTI test set	5.23
Exp. 2 (sup)	Chairs w. labels	KITTI w. labels	KITTI test set	4.43
Exp. 3 (sup)	Chairs w. labels + KITTI w. labels	N.A.	KITTI test set	4.69
Exp. 4 (unsup)	KITTI raw w/o. labels	N.A.	KITTI test set	3.67
Exp. 5 (sup)	Chairs w. labels + KITTI w. predicted labels	N.A.	KITTI test set	2.21
Exp. 6 (sup)	Chairs w. labels + KITTI w. predicted labels	KITTI w. labels	KITTI test set	2.28
Exp. 7 (sup)	Chair w. labels + KITTI w. predicted labels + KITTI w. labels	N.A.	KITTI test set	2.31

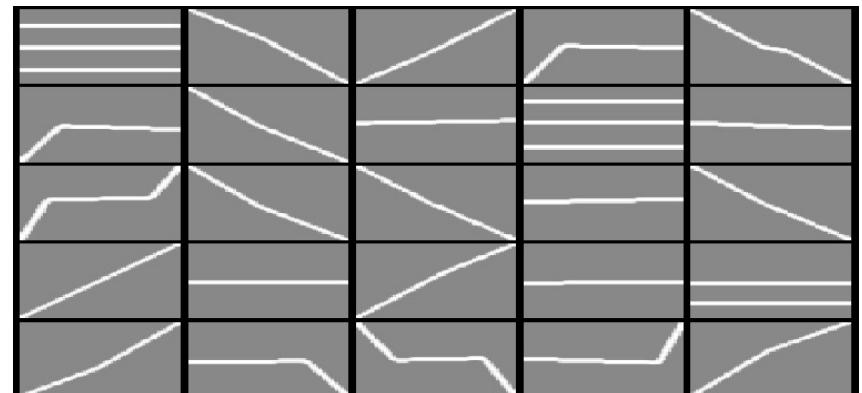
- ❖ Motivations and Importance of the problem.
- ❖ Introduction and novel contributions.
- ❖ Data synthesis in data space.
- ❖ **Learning from synthetic data.**
- ❖ Eliminating synthetic gap.
- ❖ Data synthesis in feature space.
- ❖ Conclusion.



A Challenging task.



Actual data



Synthetic data



Previous work.

- ❖ They all treat synthetic data as actual ones.[68][136][114][138][139]

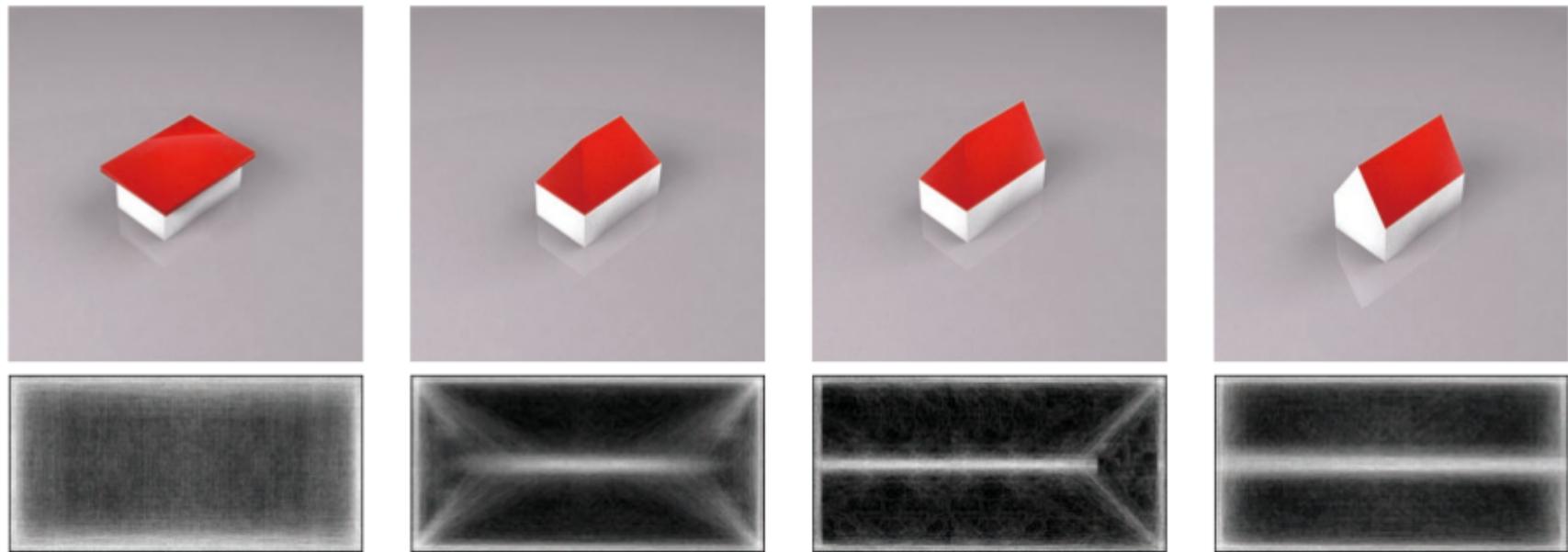
The proposed approaches.

Features has to capture almost equal amount of information from both actual and synthetic dataset.

Two types of features

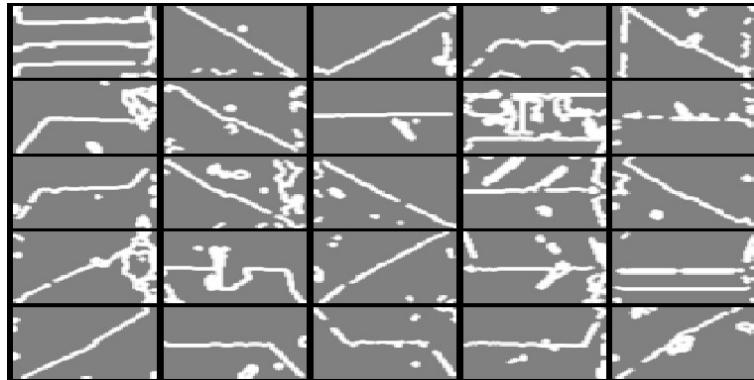
- ❖ Type I: Extract features that describe only essential structure o the data.
- ❖ Type II: Compensate additional information for synthetic data.

Showcase of Type I features

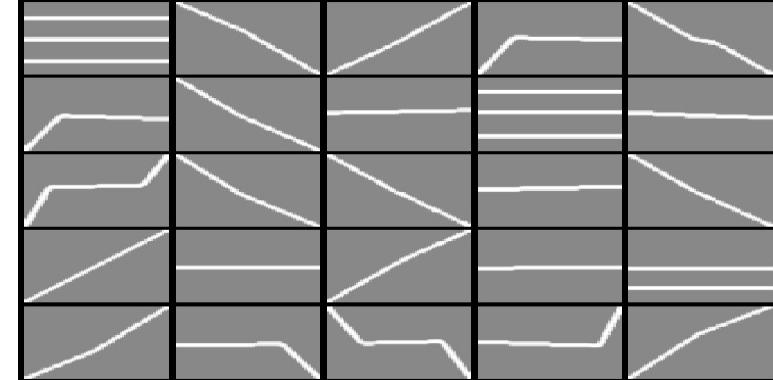


Roof type classification from satellite images [21]

- ❖ Satellite image roof style classification.



Actual data

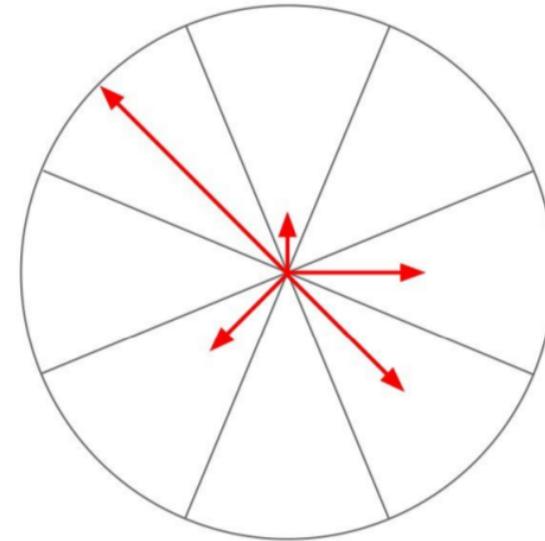
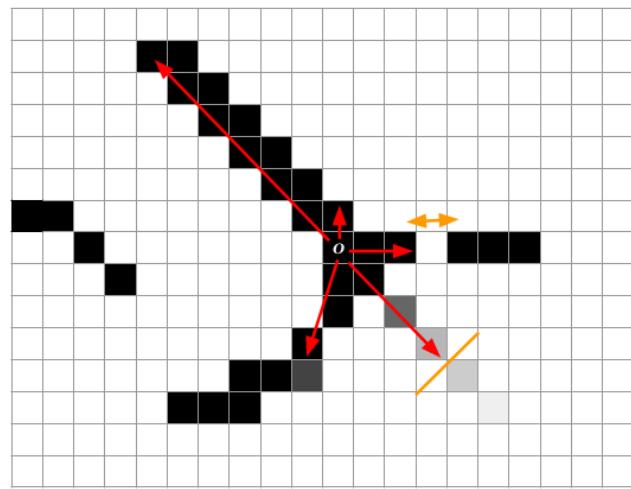


Synthetic data

Features expected:

- ❖ Ignore small blobs in actual data.
- ❖ Highlight the most evident structure of roofs.

- ❖ We proposed a feature called Histogram of Ray (HOR) in [21] to highlight edge length and direction.
- ❖ Translation, scale, rotation invariant.

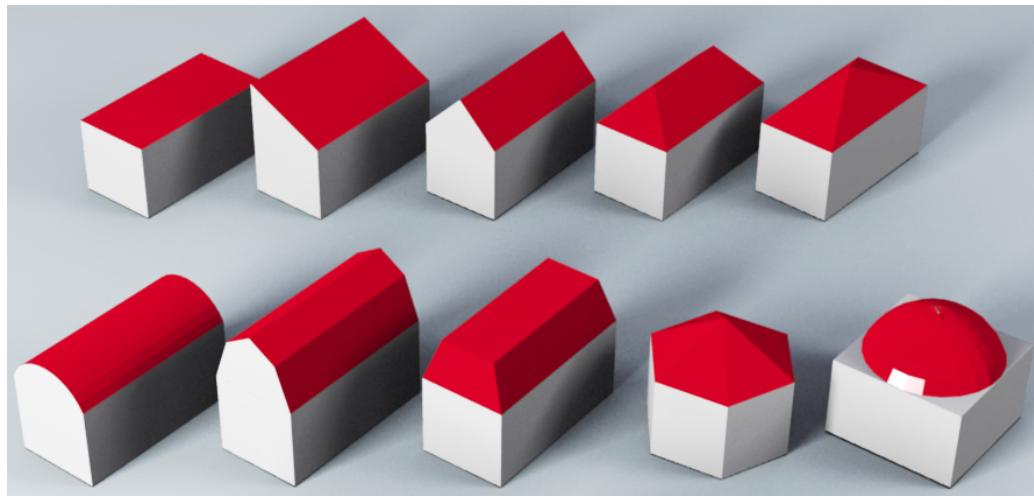


- ❖ Compare with several well-known image features:

	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

F1 score of classification results

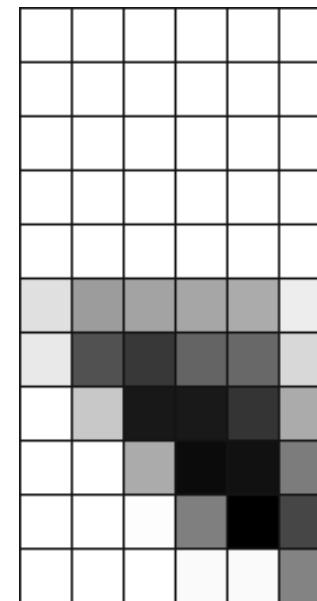
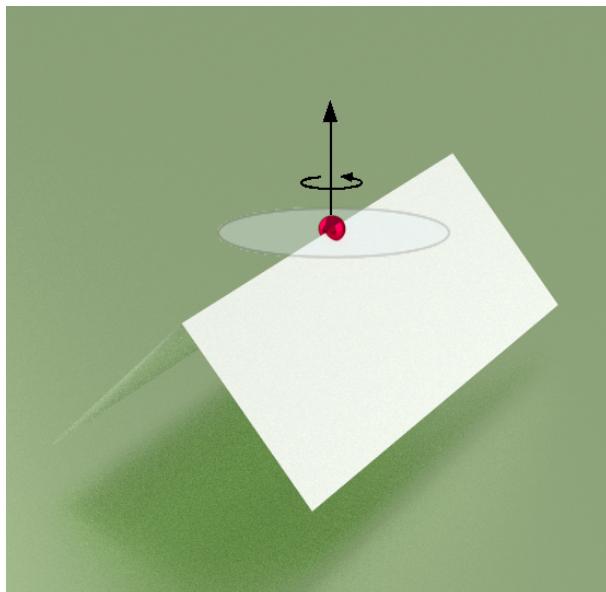
Showcase of Type II features



Roof type classification from point cloud



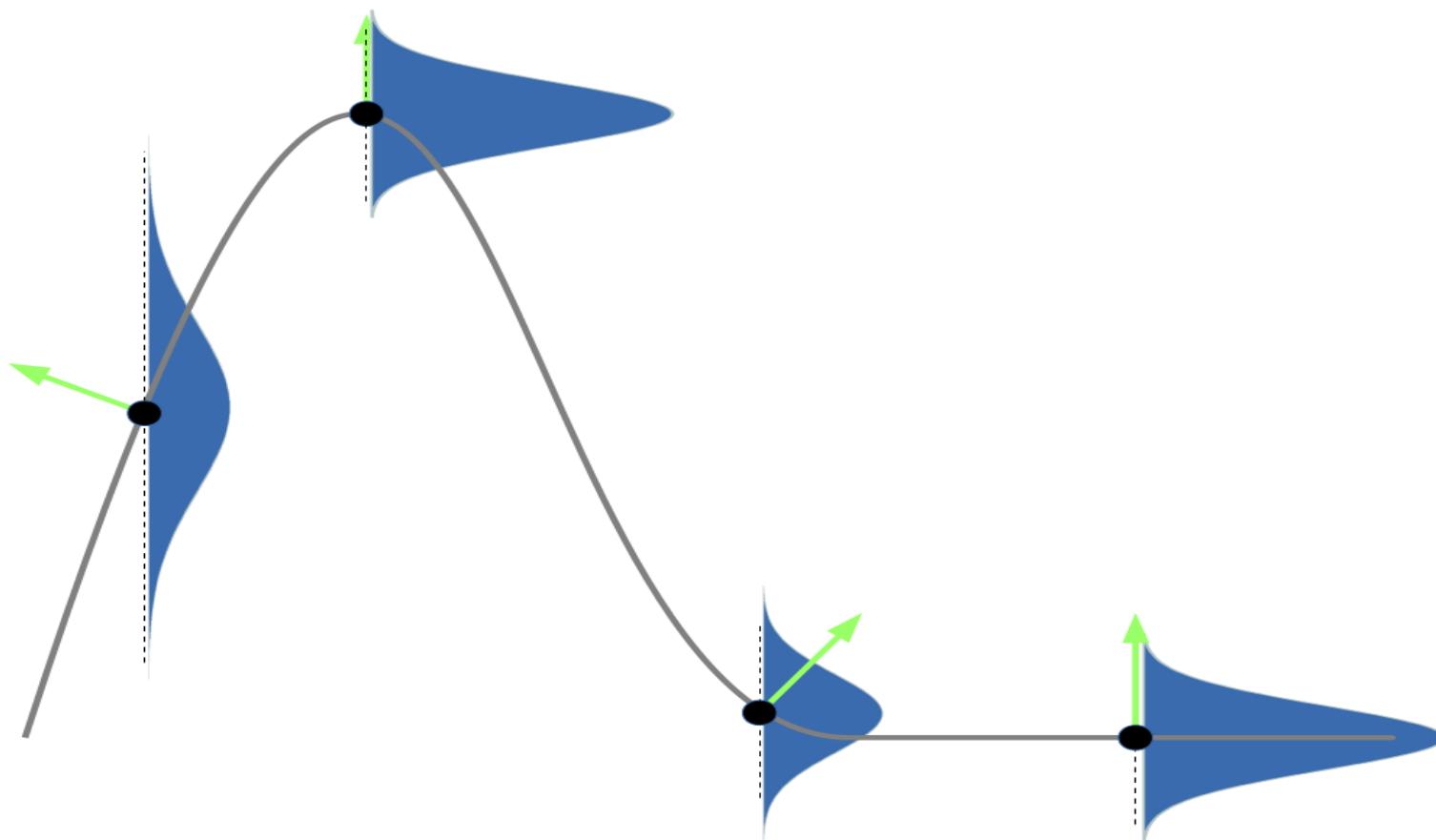
- ❖ Extract features that characterize local geometry.



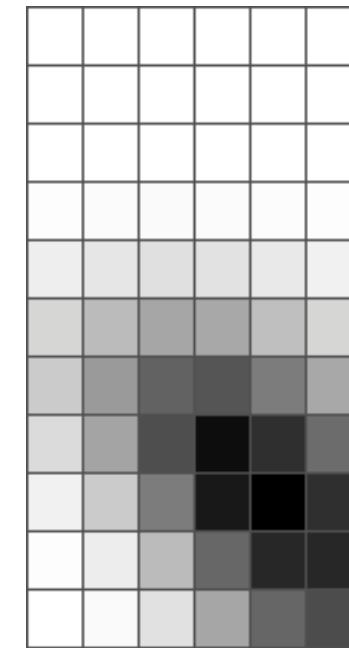
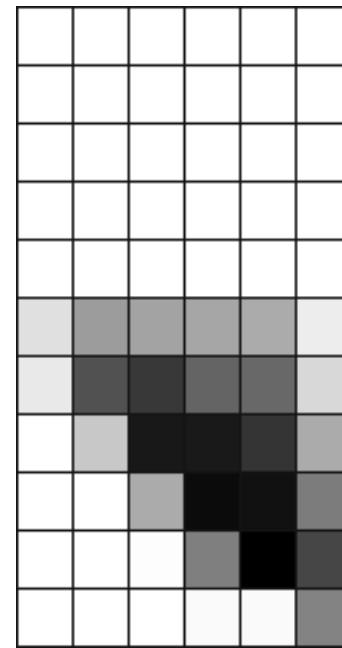
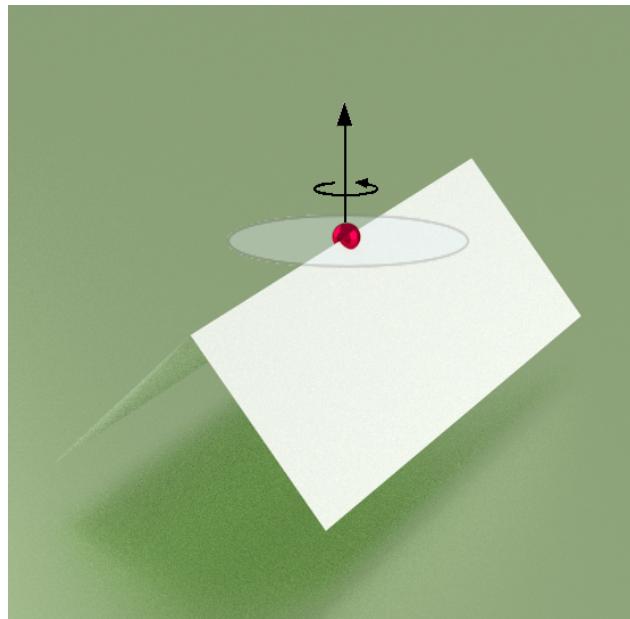
Spin image



- ❖ Learning bumpiness as a function of surface slope.



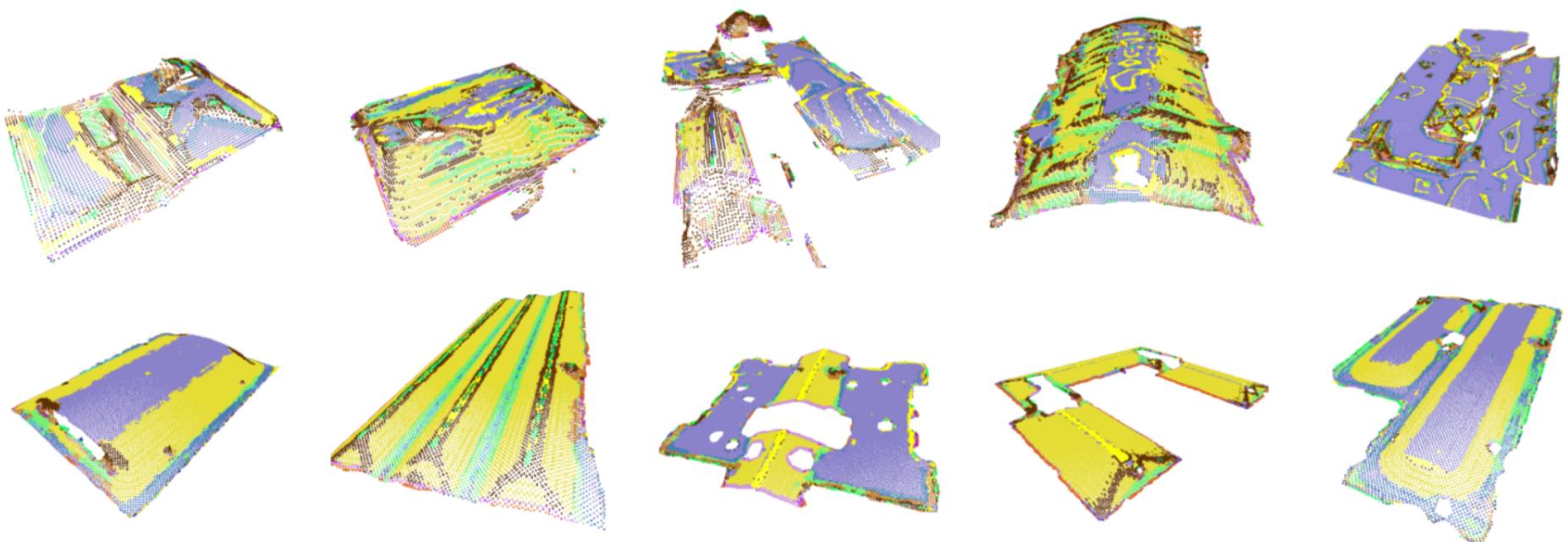
- ❖ Add random noise to synthetic data using knowledge learned.



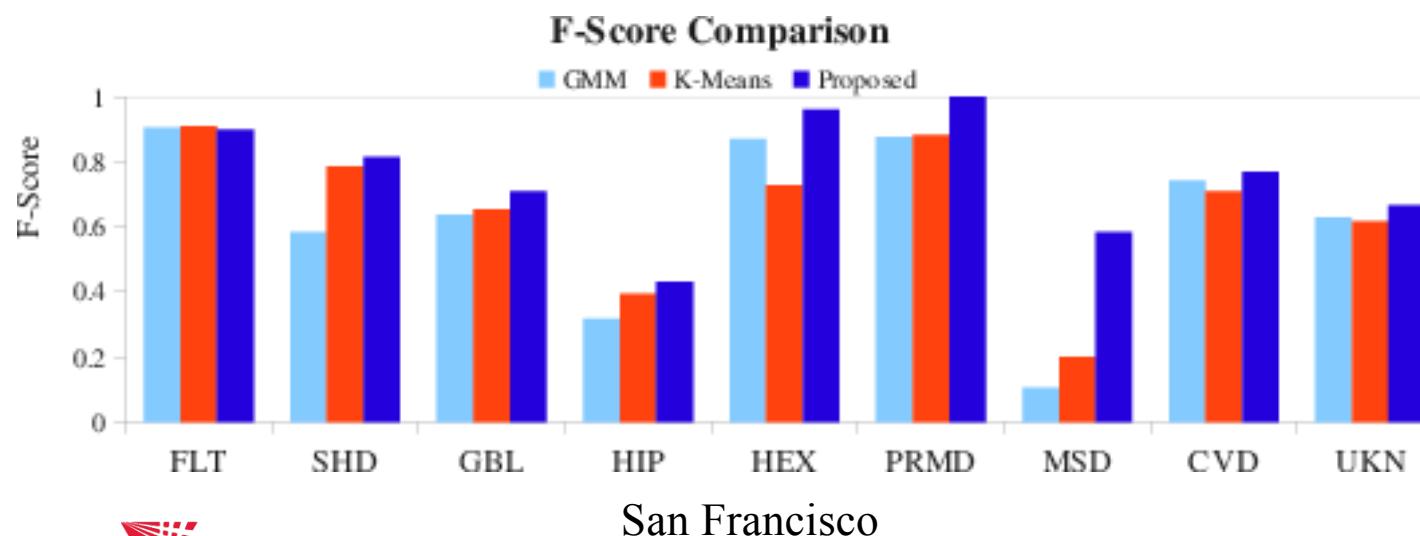
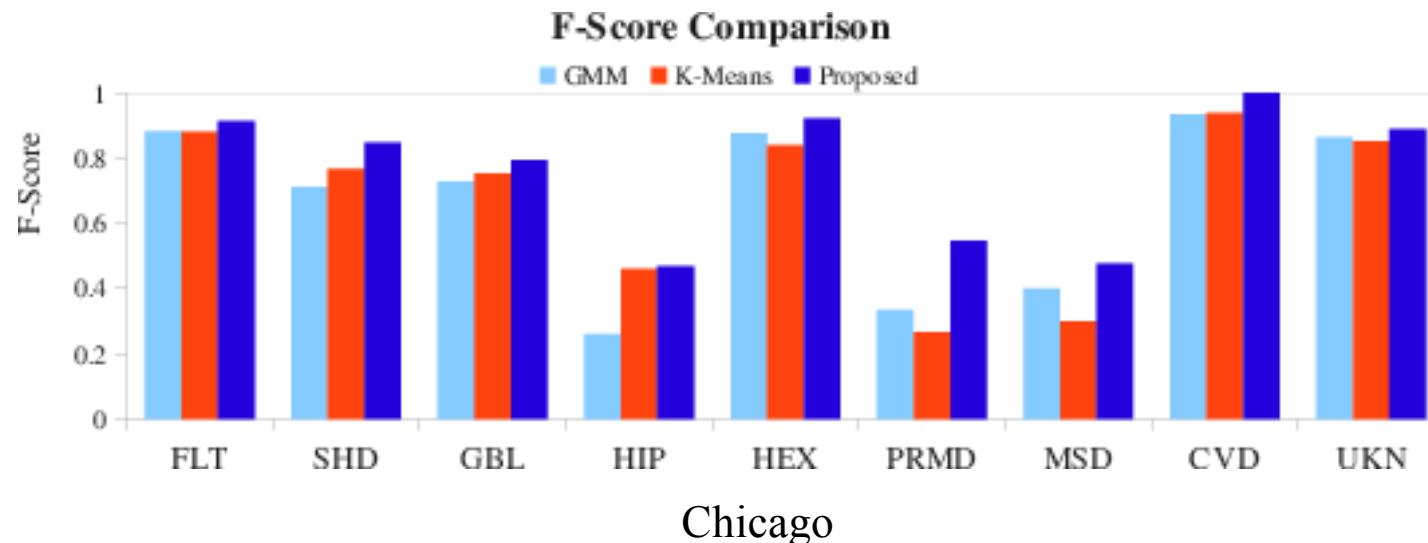
Spin image



- ❖ Point semantics classification results.



- ❖ Boost recognition rate by adding synthetic data.

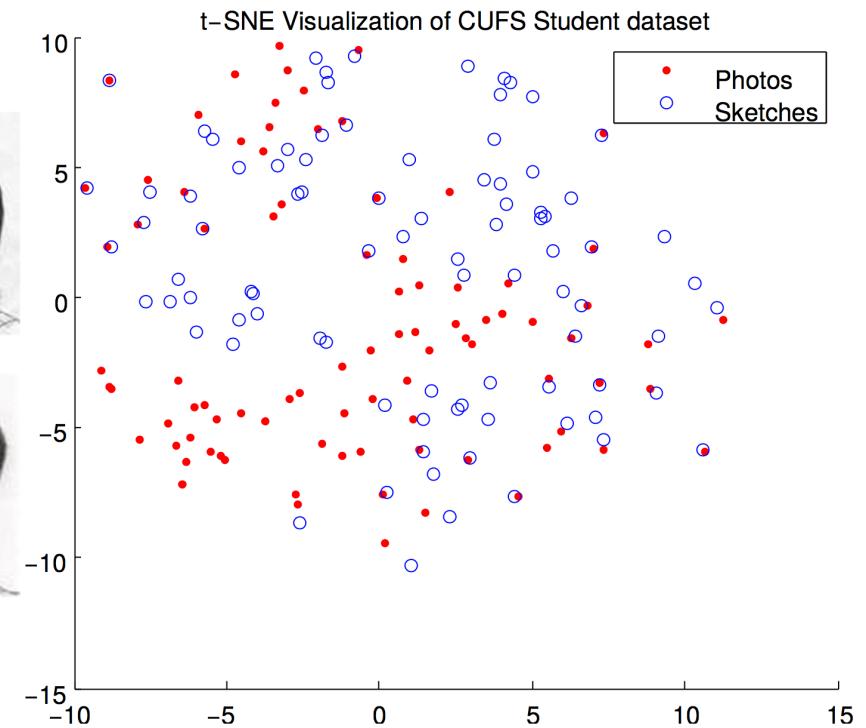


- ❖ Motivations and Importance of the problem.
- ❖ Introduction and novel contributions.
- ❖ Data synthesis in data space.
- ❖ Learning from synthetic data.
- ❖ **Eliminating synthetic gap.**
- ❖ Data synthesis in feature space.

.Conclusion.



Eliminate Synthetic Gap



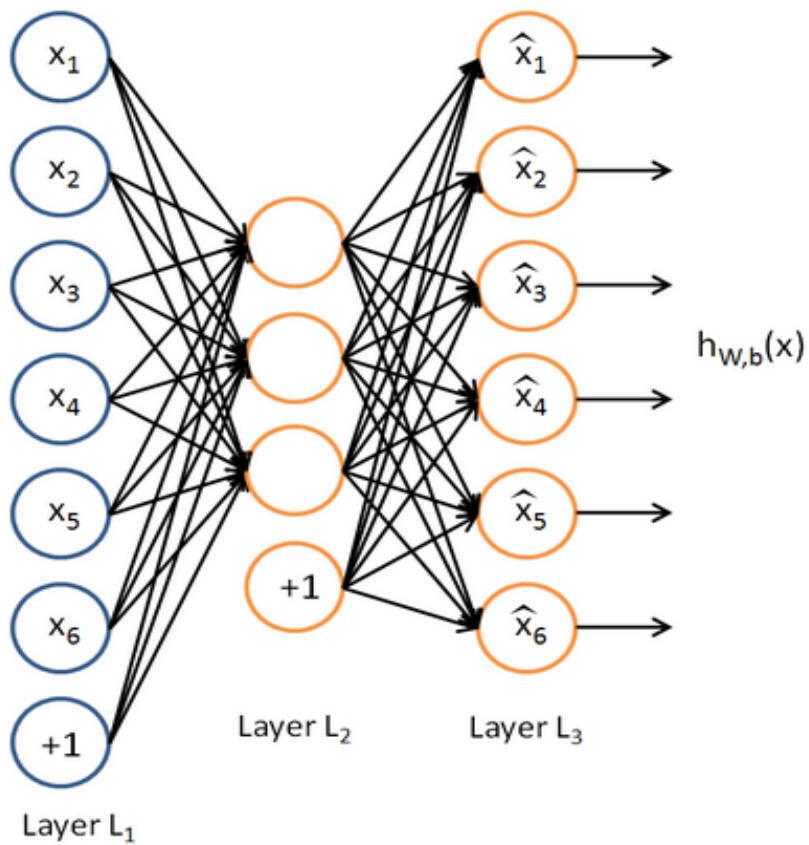
Previous work.

- ❖ 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]

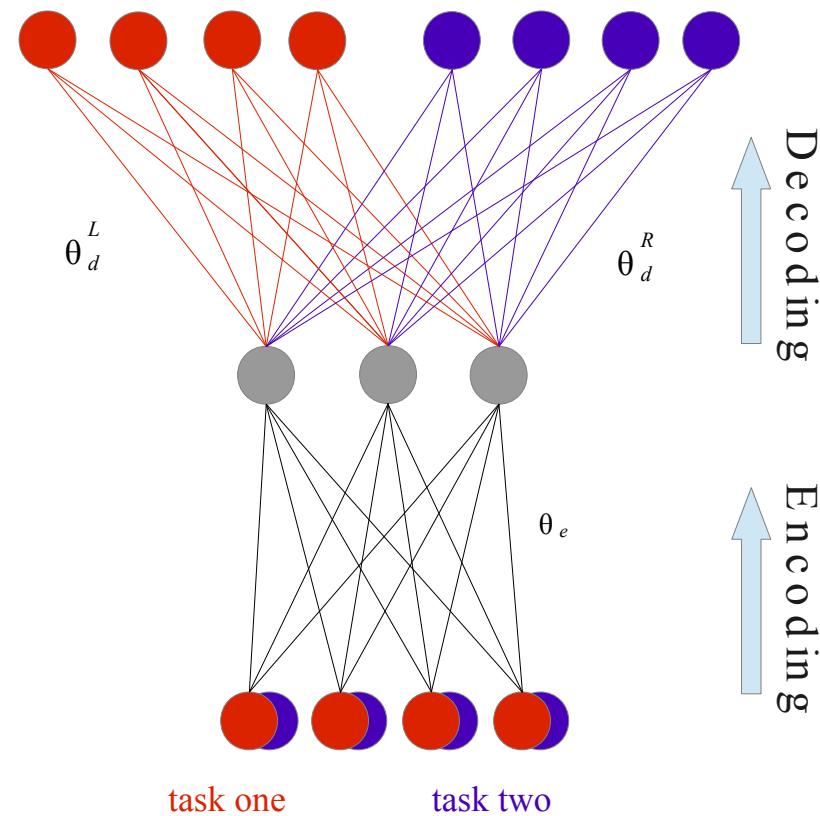
- ❖ I proposed Multi-Channel Autoencoder (MCAE)



- ❖ I proposed Multi-Channel Autoencoder (MCAE)

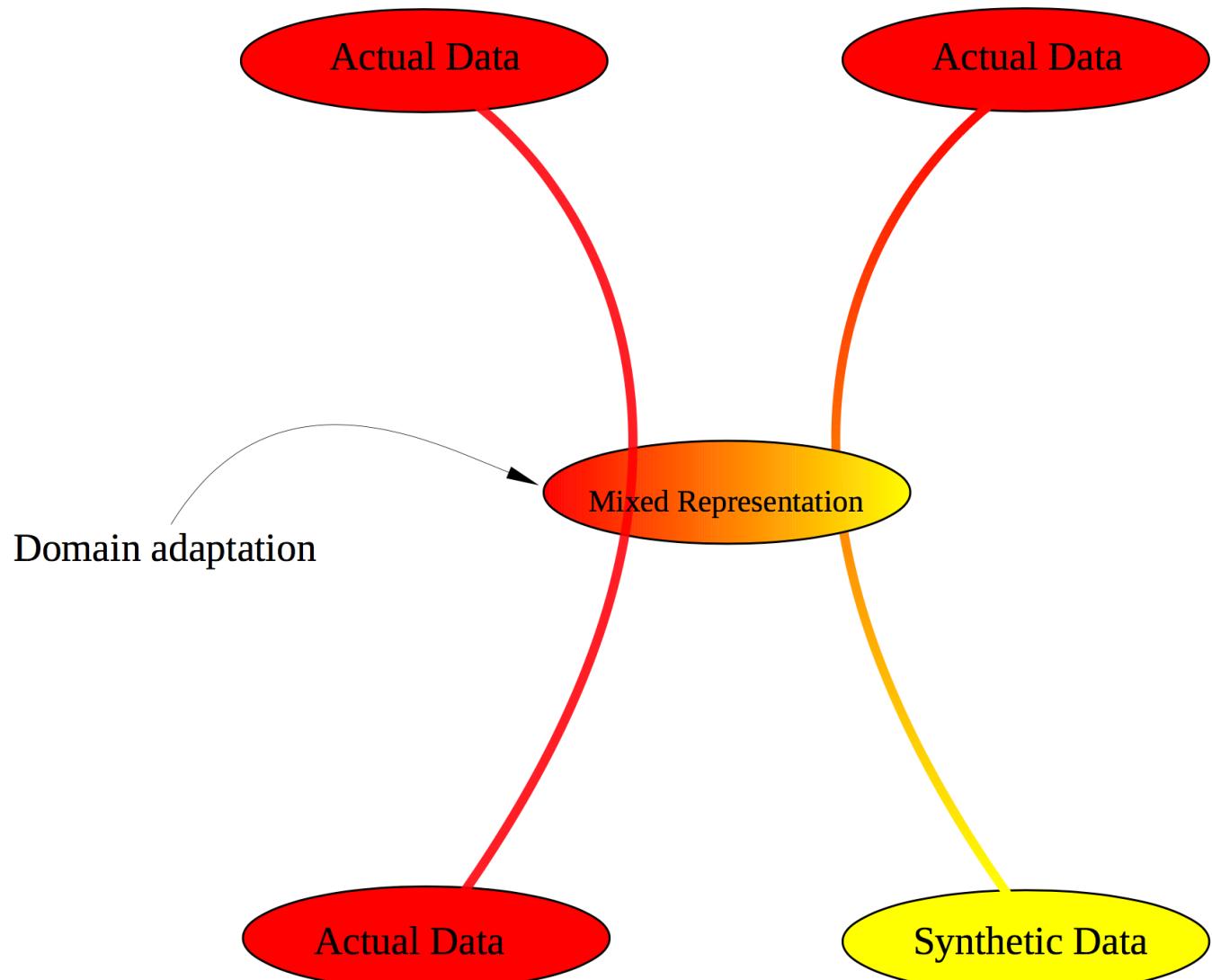


Traditional AutoEncoder



MCAE

- ❖ Configuration of MCAE



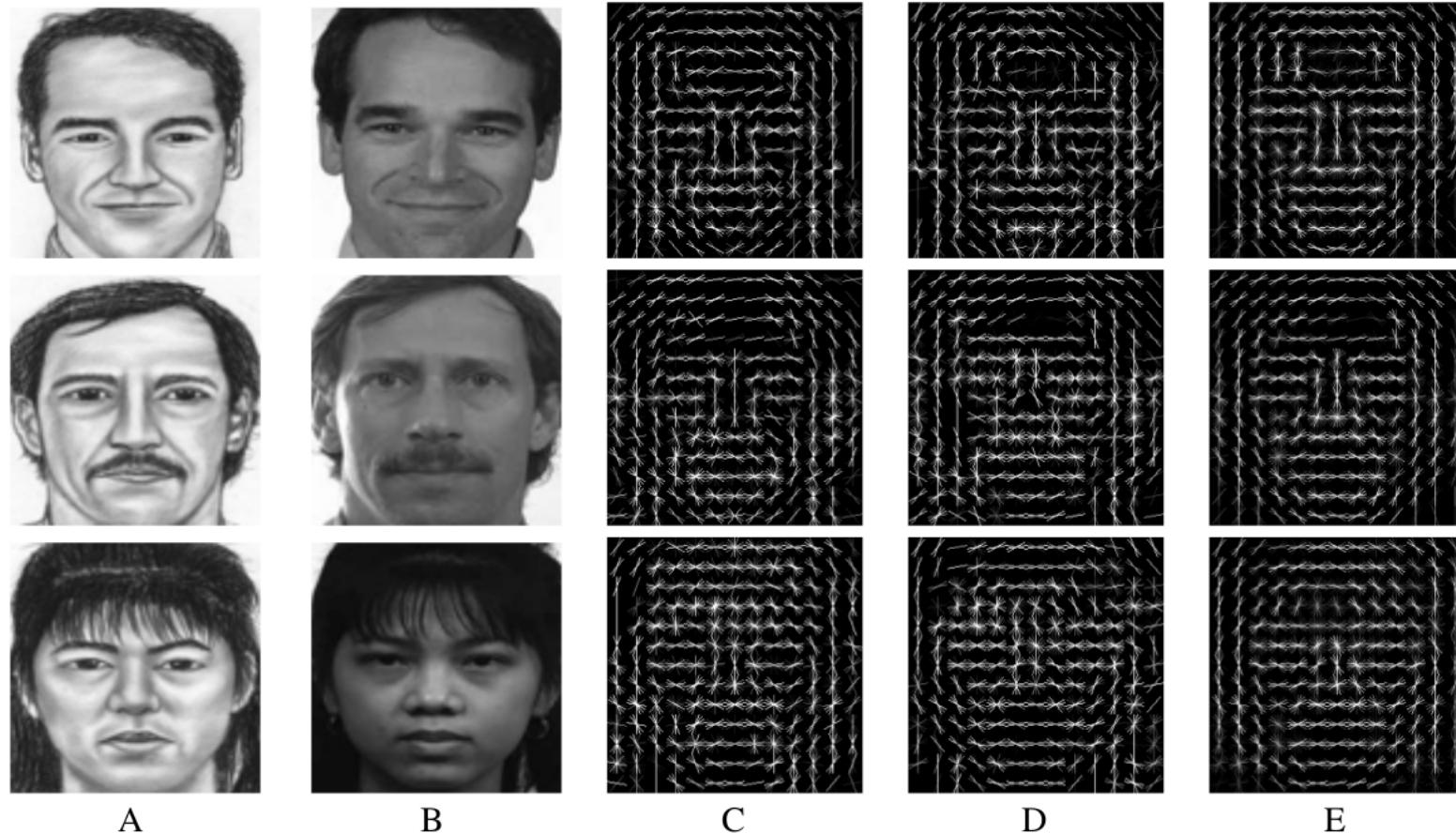
- ❖ MCAE
- ❖ Jointly learn two tasks, left and right, together.

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

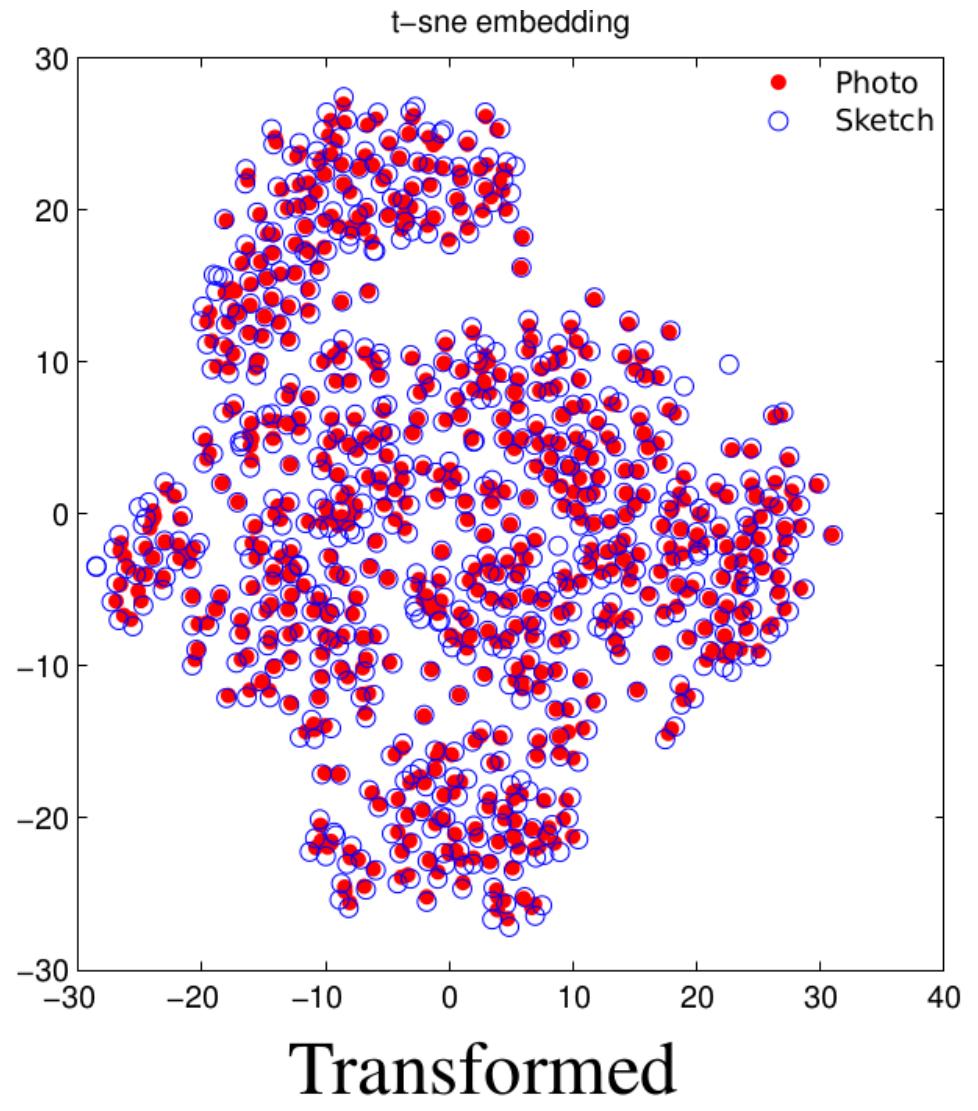
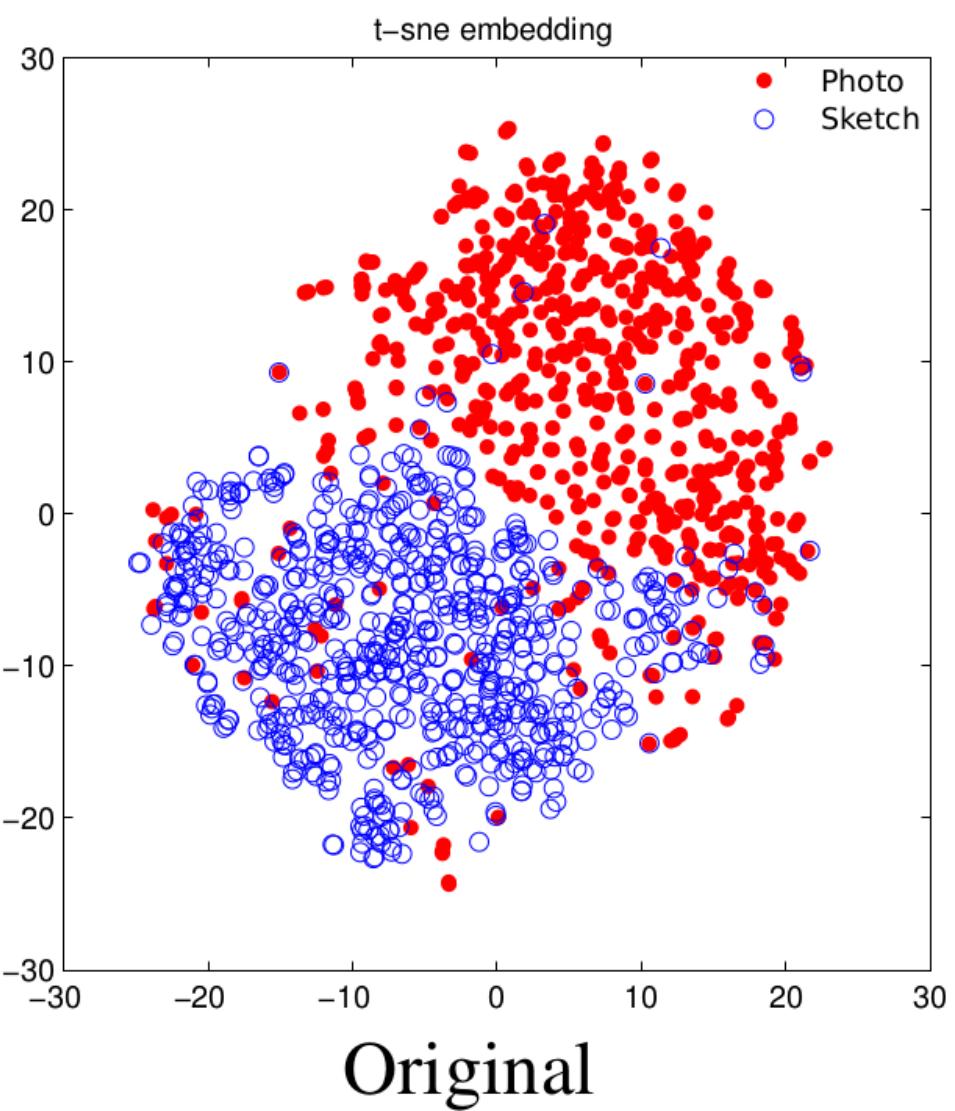
where

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

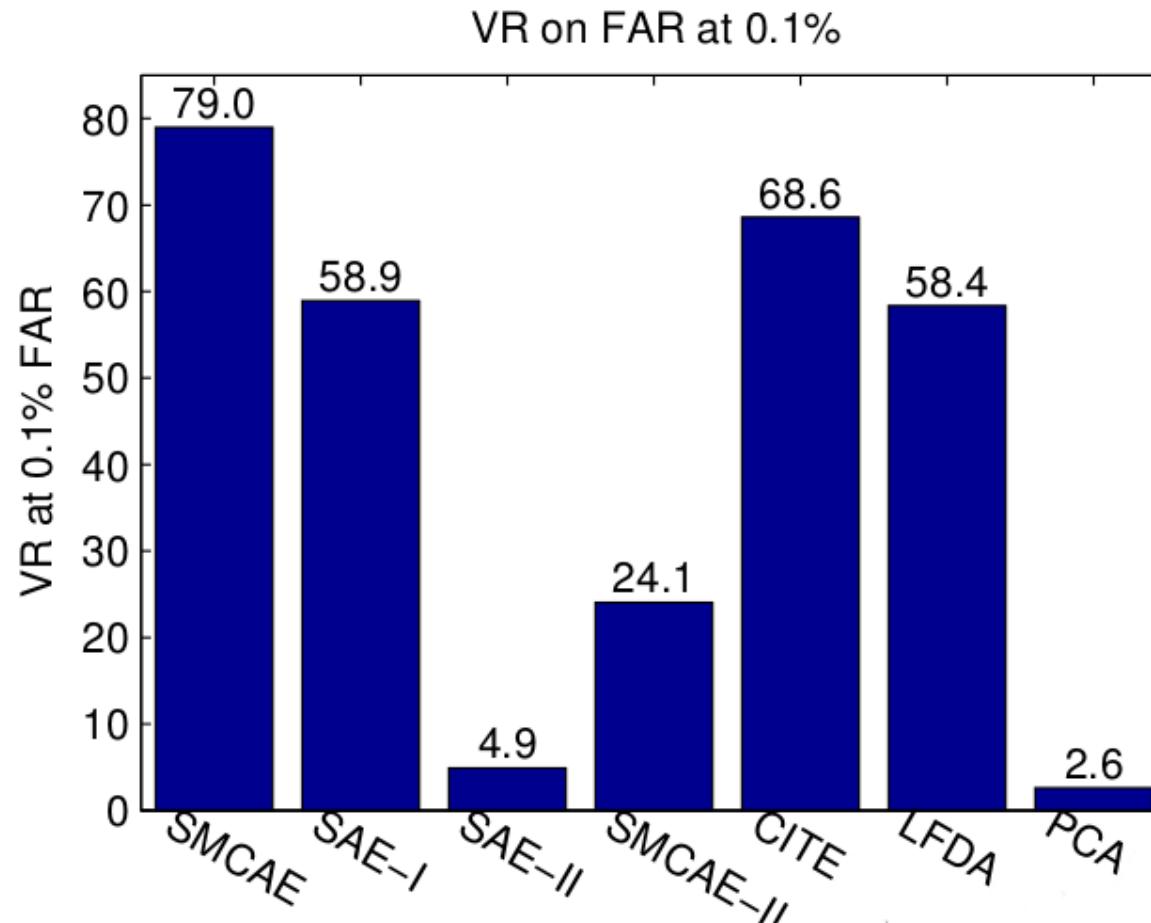
- ❖ Visualization of domain adaptation. [25]



A: Sketch. B: Photo. C: Original sketch HOG. D: Photo HOG. E: Transformed sketch HOG.



❖ Classification results



Results are evaluated on CUFSF data from [119]

- ❖ 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

The fundamental issue in imbalanced learning is the ability of imbalanced data to compromise the performance of classification algorithm.

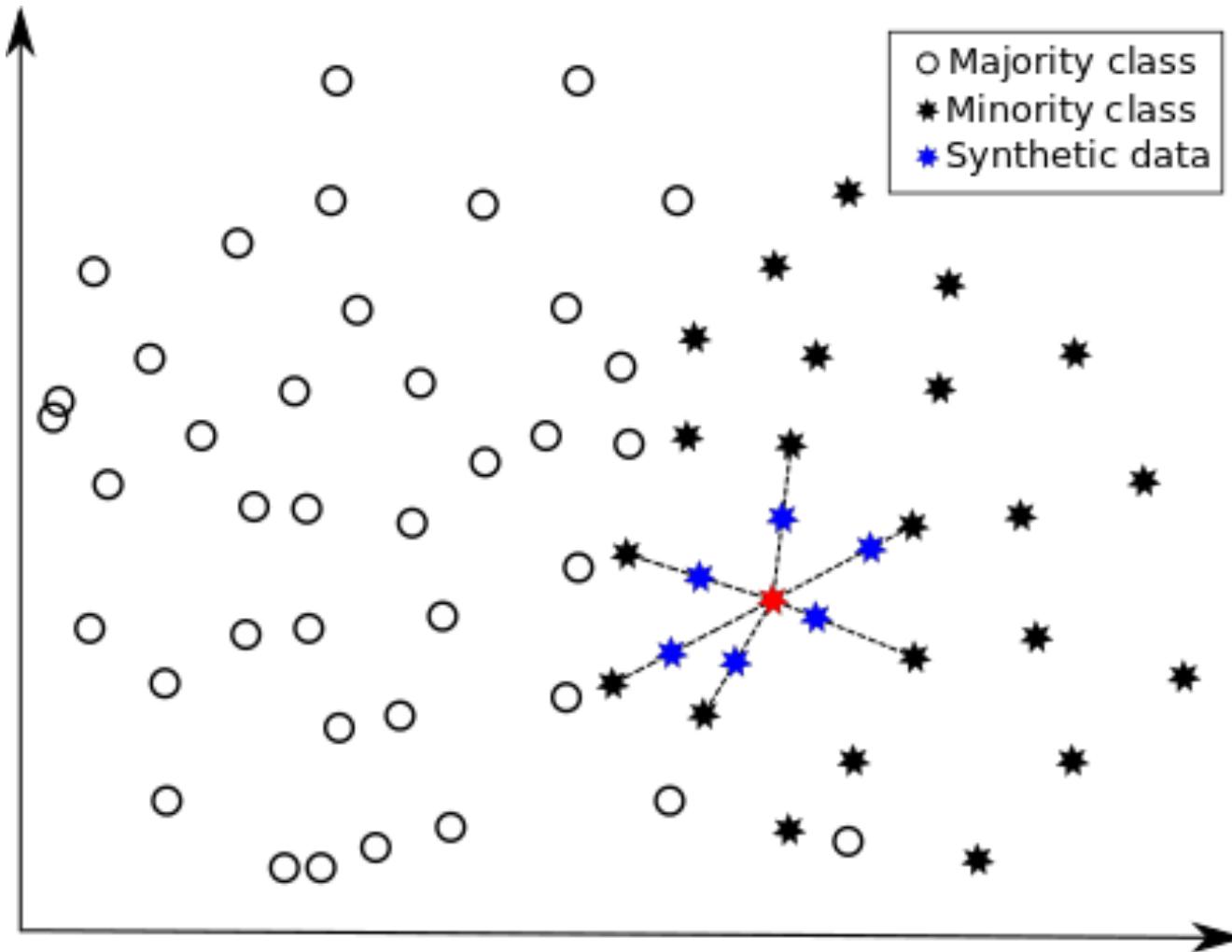
- ❖ It is hard to detect patterns, regularities within a minority class
- ❖ The general bias used in many classification algorithm make it hard to learn from the minority class.
- ❖ Noise impacts more to the minority class.

Existing work

There are primarily three groups of methods that can solve imbalanced learning problems [19].

- ❖ Cost sensitive methods.
- ❖ Kernel methods.
- ❖ Sampling methods. ★

SMOTE [4] and its variants [2][7][17][18] etc.



The proposed approach – CGMOS [27]

What is CGMOS

- ❖ CGMOS is an oversampling technique that uses the same framework proposed by SMOTE.
- ❖ CGMOS can synthesize new samples that will improve the overall *certainty* of the entire dataset in classification.
- ❖ CGMOS is theoretically proved to work better than SMOTE during training when using Bayesian classification.

Definition of certainty

Given Bayes rule in a binary classification problem:

$$P(l|x_j) = \frac{P(x_j|l)P(l)}{P(x_j)}; \quad l \in \{l_{\text{mjr}}, l_{\text{mnr}}\}$$

For a data sample (x_j, y_j) representing features and class label. The certainty for this sample in the majority and minority class are respectively defined as:

$$C(y_j = l_{\text{mjr}}|x_j) = P(y_j = l_{\text{mjr}}|x_j)$$

$$C(y_j = l_{\text{mnr}}|x_j) = P(y_j = l_{\text{mnr}}|x_j)$$

Strategy of choosing sampling seeds

- ❖ CGMOS selects a seed according to a weight $w(x_i)$ assigned to the seed.
- ❖ The weight $w(x_i)$ is computed as a *relative certainty change* comparing the certainty before and after a new sample is added.
- ❖ Given relative certainty change for sample (x_j, y_j) by adding a new sample at location x_i defined as:

$$R_{+i}(y_j|x_j) = \frac{C_{+i}(y_j|x_j) - C(y_j|x_j)}{C(y_j|x_j)}$$

Weight is computed as:

$$W(x_i) = 1 + \frac{1}{n} \sum_{j=1}^n R_{+i}(y_j|x_j)$$

Theoretical guarantee over SMOTE

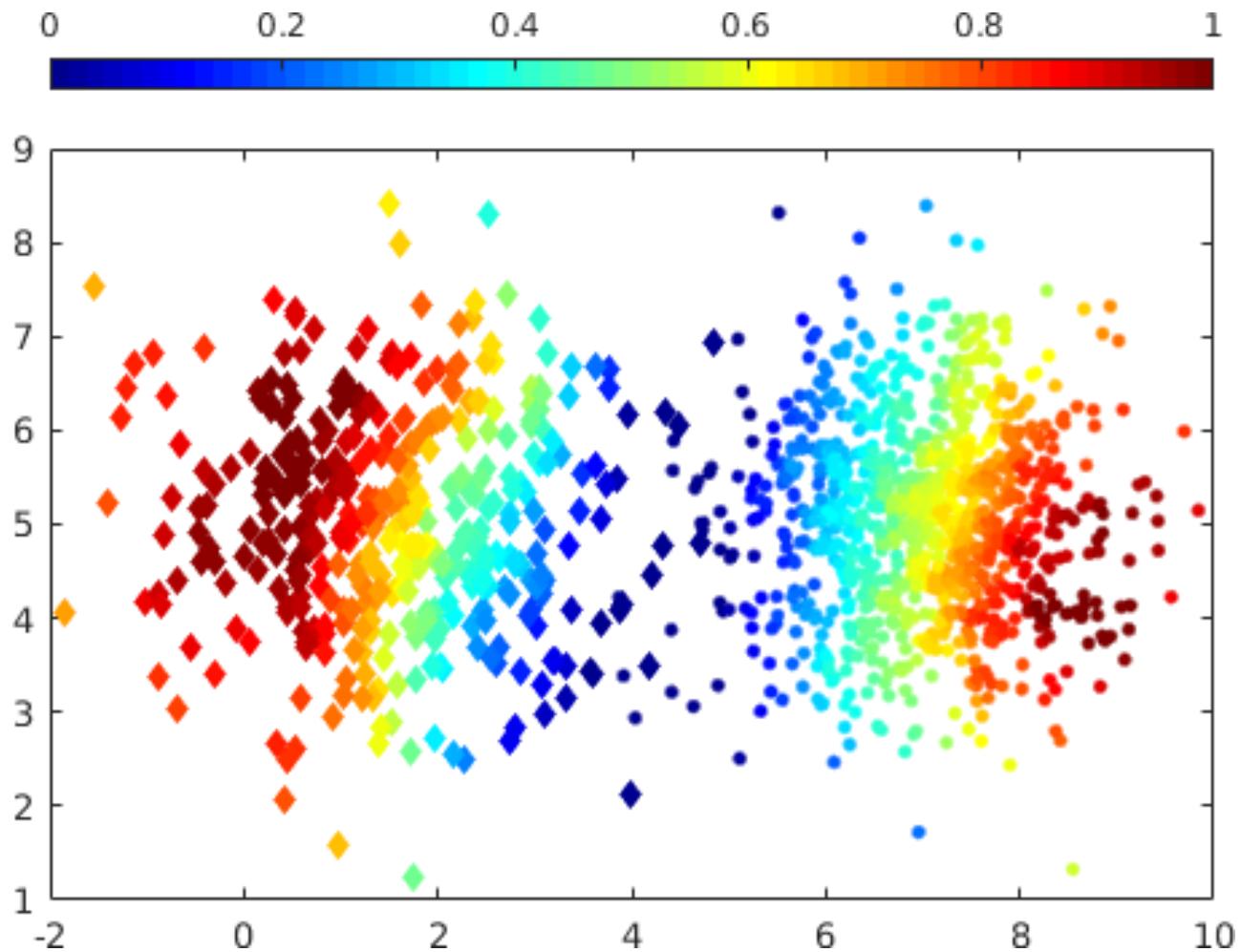
Definition (Average gain): The average gain when adding sample x_i is defined by:

$$\bar{r}_{+i} = \frac{1}{n} \sum_{j=1}^n r_{+i}(y_j|x_j)$$

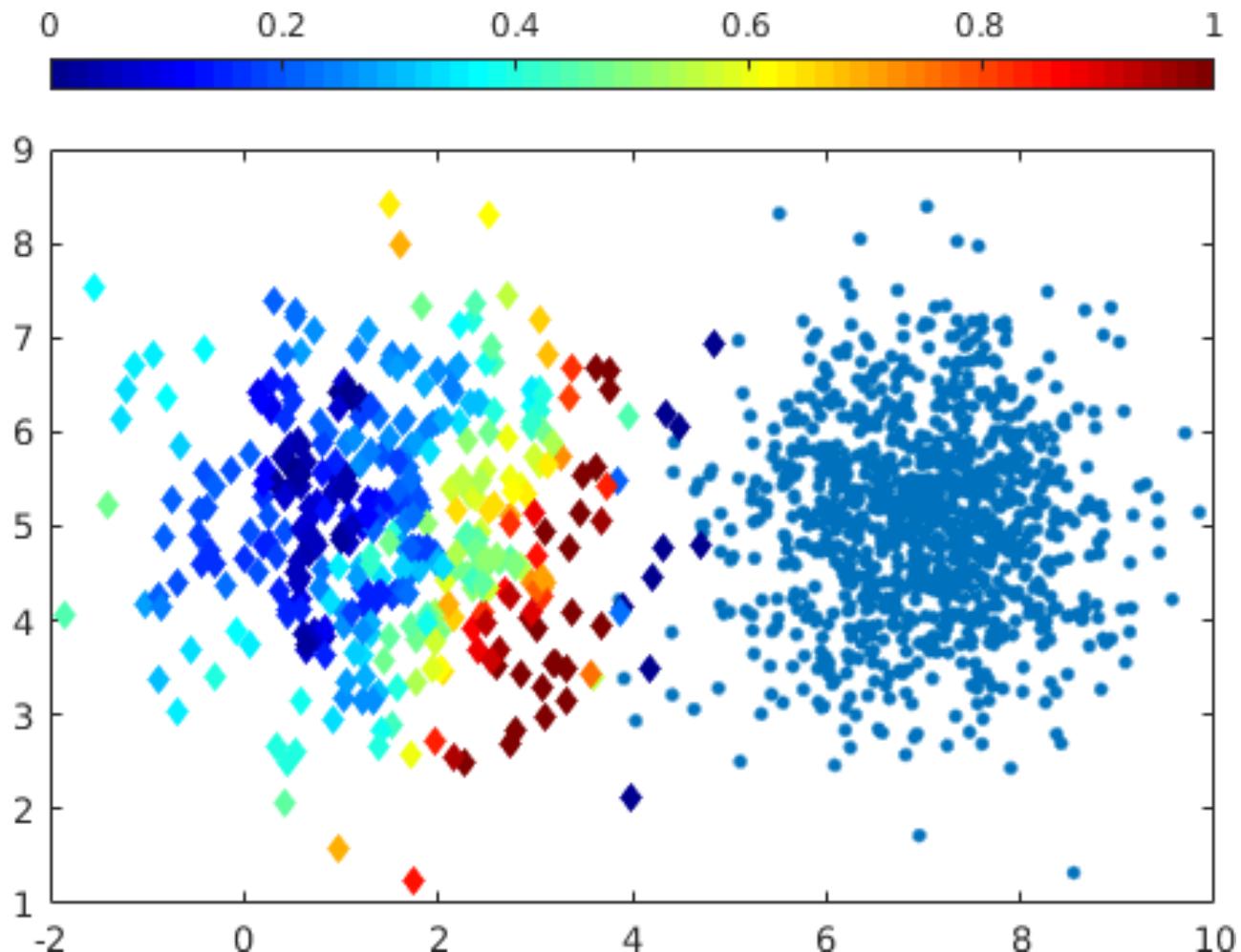
where

$$r_{+i}(y_j|x_j) = 1 + R_{+i}(y_j|x_j).$$

Theorem: *The expected average gain in CGMOS is higher or equal to that of SMOTE.*



Certainty



Relative certainty change

Results

Datasets: 30 datasets downloaded from UC Irvine machine learning repository corresponding to existing evaluations [160].

Compared approaches: No addition, Duplication, SMOTE [35], Boarderline-SMOTE [36], ADASYN [37], MWMOTE [156], and RAMOBoost [154].

Classifiers: Bayes with KDE, K nearest neighbors, Adaboost.M1, SVM, Neural networks, Random forest according to existing evaluations.

	CGMOS	Original	Dup	SMOTE	B-SMOTE	MWMOTE	ADASYN	RAMOboost
BankMarket	0.728	0.661	0.708	0.718	0.710	0.721	0.710	0.723
BloodService	0.733	0.653	0.648	0.649	0.651	0.720	0.714	0.728
BreastCancer	0.992	0.992	0.993	0.992	0.989	0.991	0.991	0.992
BreastTissue	0.984	0.899	0.946	0.932	0.917	0.937	0.908	0.943
CarEvaluation	0.997	0.995	0.845	0.997	0.994	0.996	0.997	0.995
Card'graphy	0.977	0.976	0.939	0.962	0.956	0.925	0.957	0.960
CharacterTraj	0.985	0.962	0.717	0.985	0.978	0.981	0.988	0.909
Chess	0.977	0.974	0.959	0.973	0.977	0.974	0.975	0.959
ClimateSim	0.908	0.908	0.861	0.902	0.863	0.901	0.901	0.882
Contraceptive	0.724	0.705	0.699	0.712	0.702	0.705	0.702	0.705
Fertility	0.673	0.615	0.594	0.634	0.592	0.604	0.639	0.638
Haberman	0.651	0.623	0.577	0.600	0.593	0.594	0.587	0.586
ILPD	0.707	0.687	0.693	0.715	0.703	0.702	0.693	0.703
ImgSeg	0.999	0.998	0.999	0.997	0.998	0.998	0.997	0.998
Leaf	0.908	0.880	0.782	0.852	0.775	0.836	0.839	0.821
Libras	0.945	0.922	0.859	0.929	0.886	0.936	0.923	0.883
MultipleFs	0.998	0.998	0.997	0.998	0.997	0.997	0.996	0.997
Parkinson	0.841	0.676	0.692	0.834	0.791	0.837	0.842	0.760
PlanRelax	0.472	0.457	0.494	0.469	0.445	0.467	0.488	0.464
QSAR	0.901	0.886	0.879	0.895	0.863	0.886	0.886	0.882
SPECT	0.820	0.772	0.803	0.808	0.811	0.752	0.801	0.799
SPECTF	0.819	0.819	0.800	0.805	0.816	0.812	0.825	0.795
SeismicBumps	0.743	0.735	0.712	0.727	0.740	0.732	0.715	0.691
Statlog	0.998	0.992	0.996	0.998	0.990	0.996	0.976	0.996
PlatesFaults	0.956	0.928	0.844	0.954	0.920	0.943	0.956	0.881
TAEvaluation	0.748	0.682	0.644	0.703	0.671	0.707	0.665	0.657
UserKnowledge	0.958	0.837	0.919	0.953	0.947	0.951	0.950	0.888
Vertebral	0.890	0.839	0.869	0.855	0.829	0.860	0.794	0.872
Customers	0.952	0.930	0.943	0.946	0.884	0.902	0.946	0.952
Yeast	0.925	0.792	0.844	0.907	0.898	0.900	0.906	0.851
Average	0.864	0.827	0.808	0.844	0.830	0.842	0.842	0.830



- ❖ 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.



Conclusion

Goals:

- Use data synthesis techniques to
- ❖ Boost performance of object recognition.
 - ❖ Ease ground truth labeling process.

Novel contributions:

- ❖ Data synthesis in data space.
- ❖ Learning from synthetic data.
- ❖ Eliminating synthetic gap.
- ❖ Data synthesis in feature space.

Future works:

- ❖ Deep learning based data synthesis.
- ❖ Automatically synthesize more complicated data such as video, 3D object, etc..



Thank you very much

Questions and Comments?

8	2	4	3	1	4	3	5	3	7
3	6	6	3	5	4	6	6	0	0
0	6	8	2	7	8	0	8	3	4
4	5	1	6	9	3	5	7	2	6
0	4	7	6	2	8	7	6	6	0
0	4	8	3	2	3	1	5	7	3
8	5	0	3	6	8	7	4	2	0
3	3	2	7	1	4	5	0	4	2
9	6	3	4	9	8	1	8	9	6
6	6	3	6	8	4	0	3	1	1

Real Data

8	2	4	3	1	4	3	5	3	7
3	6	0	3	5	4	6	6	0	0
0	6	4	2	7	8	0	8	3	4
4	5	1	6	9	3	5	7	2	6
0	4	7	6	2	8	7	6	6	0
0	4	8	3	2	3	1	5	7	3
8	5	0	3	6	8	7	4	2	0
3	3	2	7	1	4	5	0	4	2
9	6	3	4	9	8	1	8	9	6
6	6	3	6	9	4	0	3	1	1

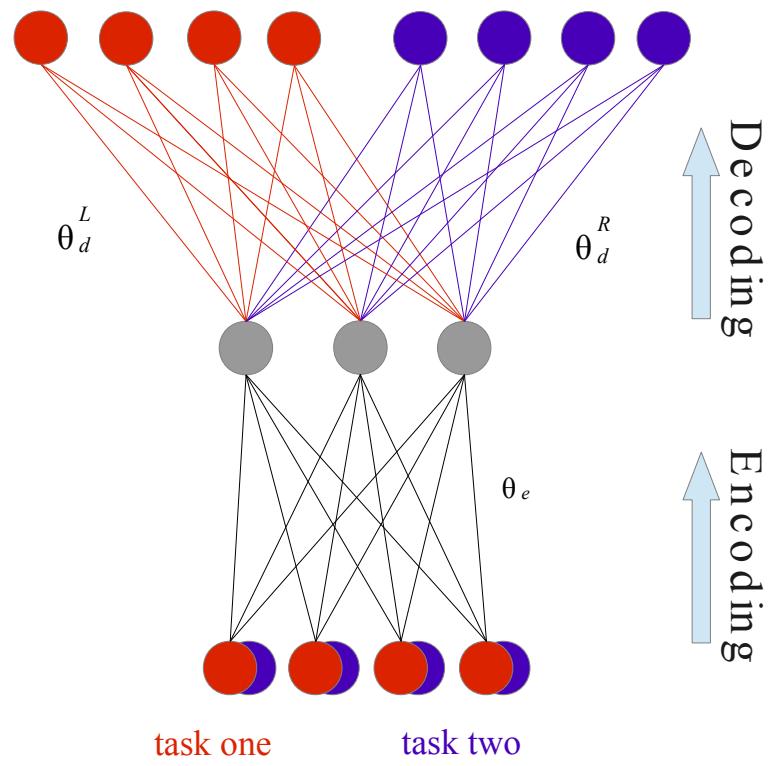
Synthesized Data

- ❖ Roof style classification results compared to unsupervised approaches.

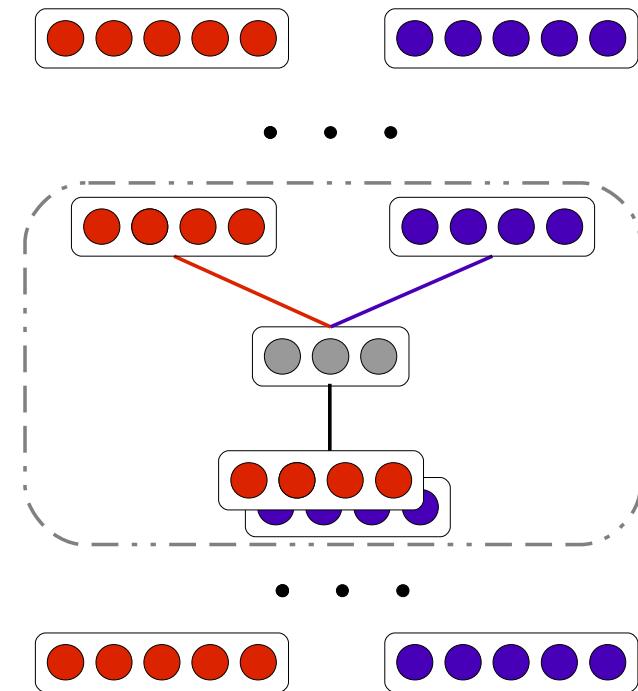
	Chicago						San Francisco					
	Precision			Recall			Precision			Recall		
	GMM	KM	Ours	GMM	KM	Ours	GMM	KM	Ours	GMM	KM	Ours
FLAT	0.87	0.85	0.92	0.88	0.90	0.90	0.88	0.88	0.86	0.93	0.93	0.93
SHED	0.94	0.94	0.91	0.57	0.64	0.78	0.87	0.91	1.00	0.43	0.68	0.68
GABLE	0.62	0.67	0.71	0.86	0.84	0.88	0.57	0.61	0.65	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



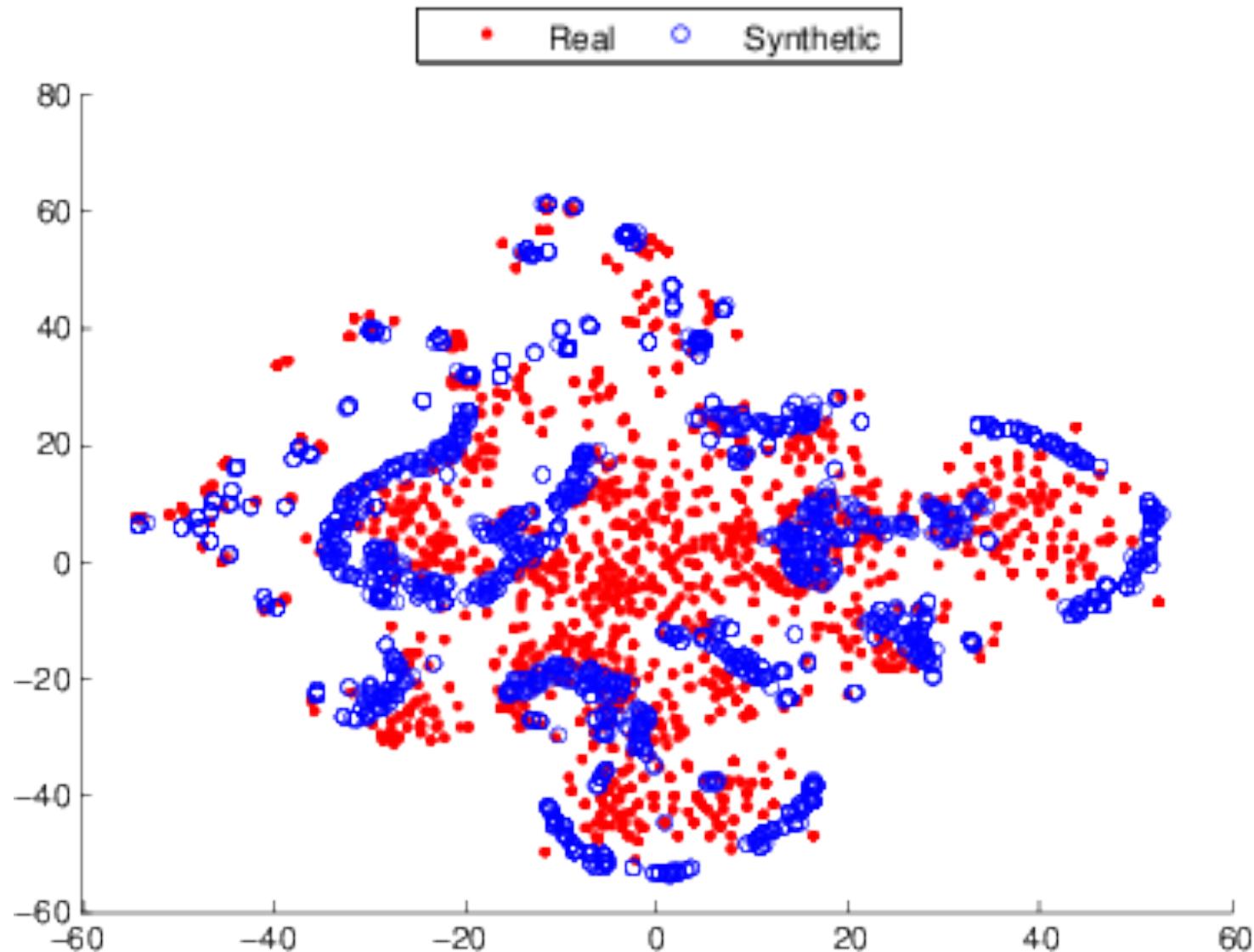
The proposed MCAE



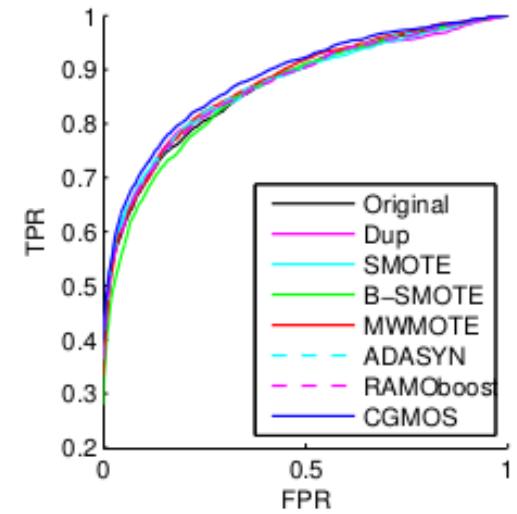
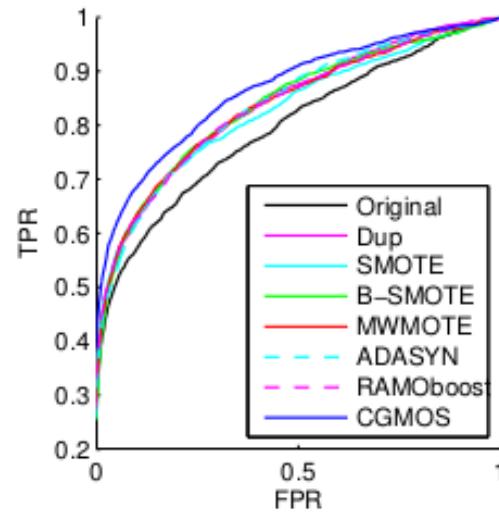
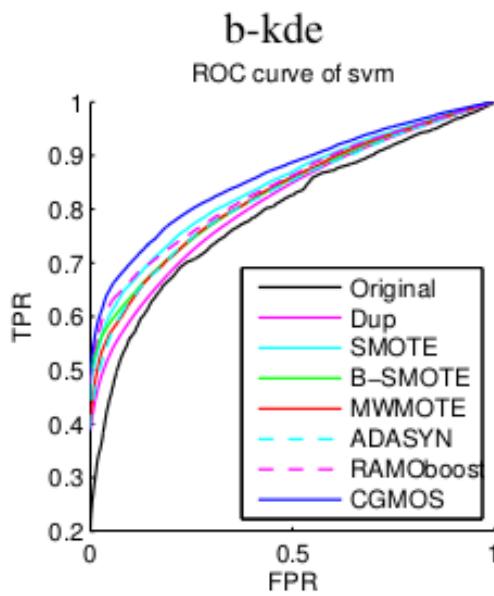
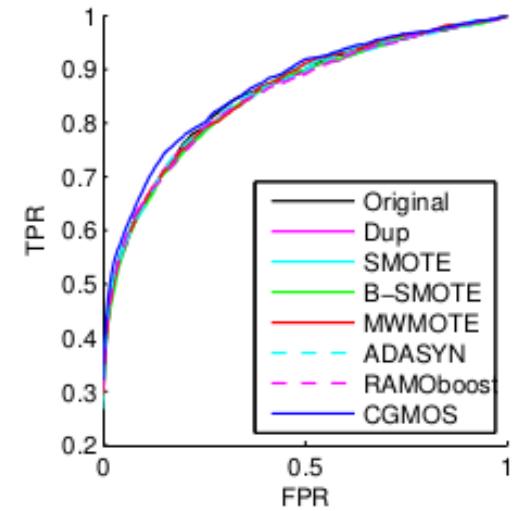
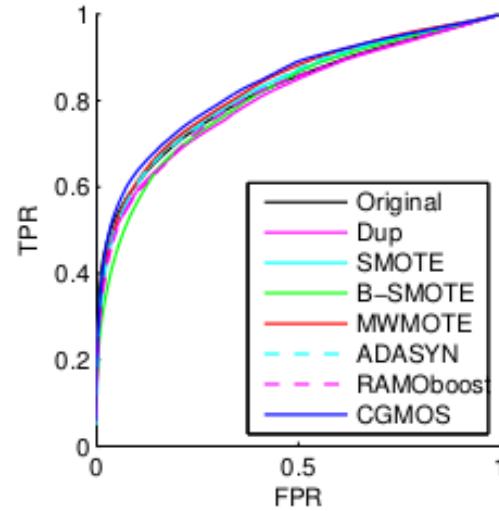
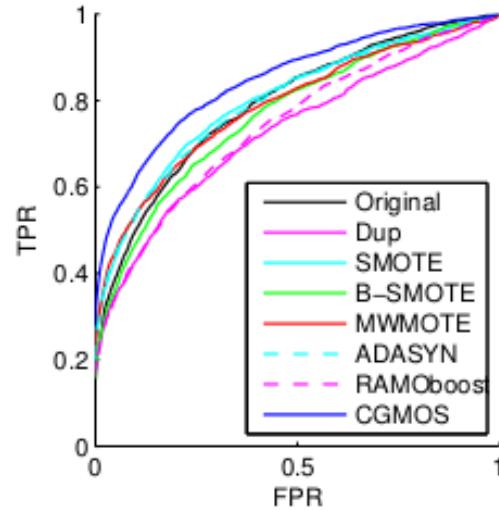
Basic structure



Stacked up

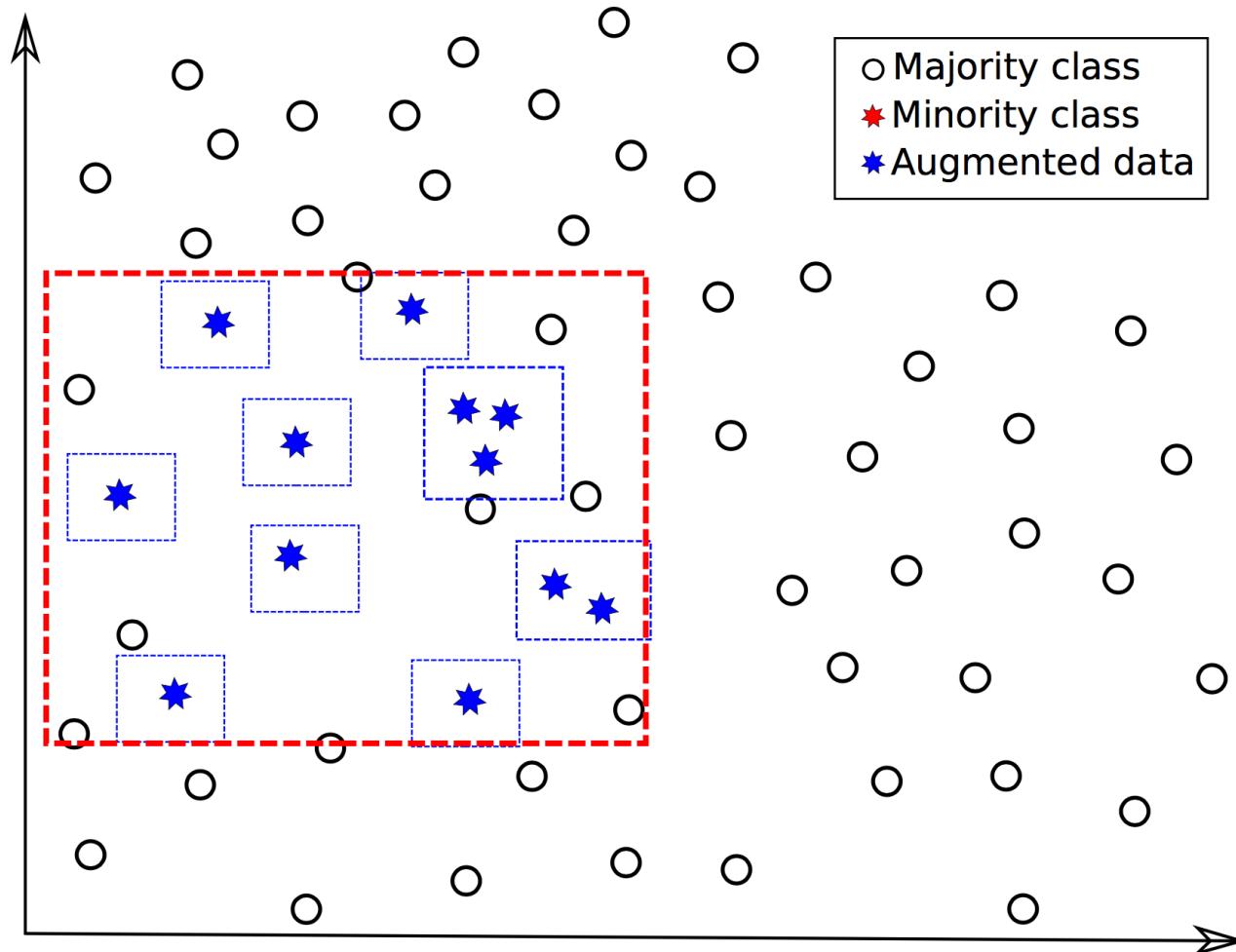


ROC curves



Existing sampling methods

1) Creating identical samples.



Statistical significance analysis

Given the significance level as 5%, we compute the p-values of statistical significance tests of classification results using CGMOS against the compared methods:

	Knn	Rf	B-kde	Nn	Svm	Boost
Original	5e-5	1e-4	0.004	1e-4	0.026	0.04
Dup	2e-6	5e-5	3e-6	0.03	0.049	0.004
SMOTE	0.003	2e-4	6e-6	0.018	0.006	0.046
B-SMOTE	4e-6	7e-6	2e-5	5e-4	0.047	5e-4
MWMOTE	0.046	4e-5	1e-5	0.003	0.005	0.007
ADASYN	8e-6	7e-5	9e-5	0.005	1e-4	0.003
RAMOboost	2e-6	5e-5	3e-6	0.001	0.045	0.035

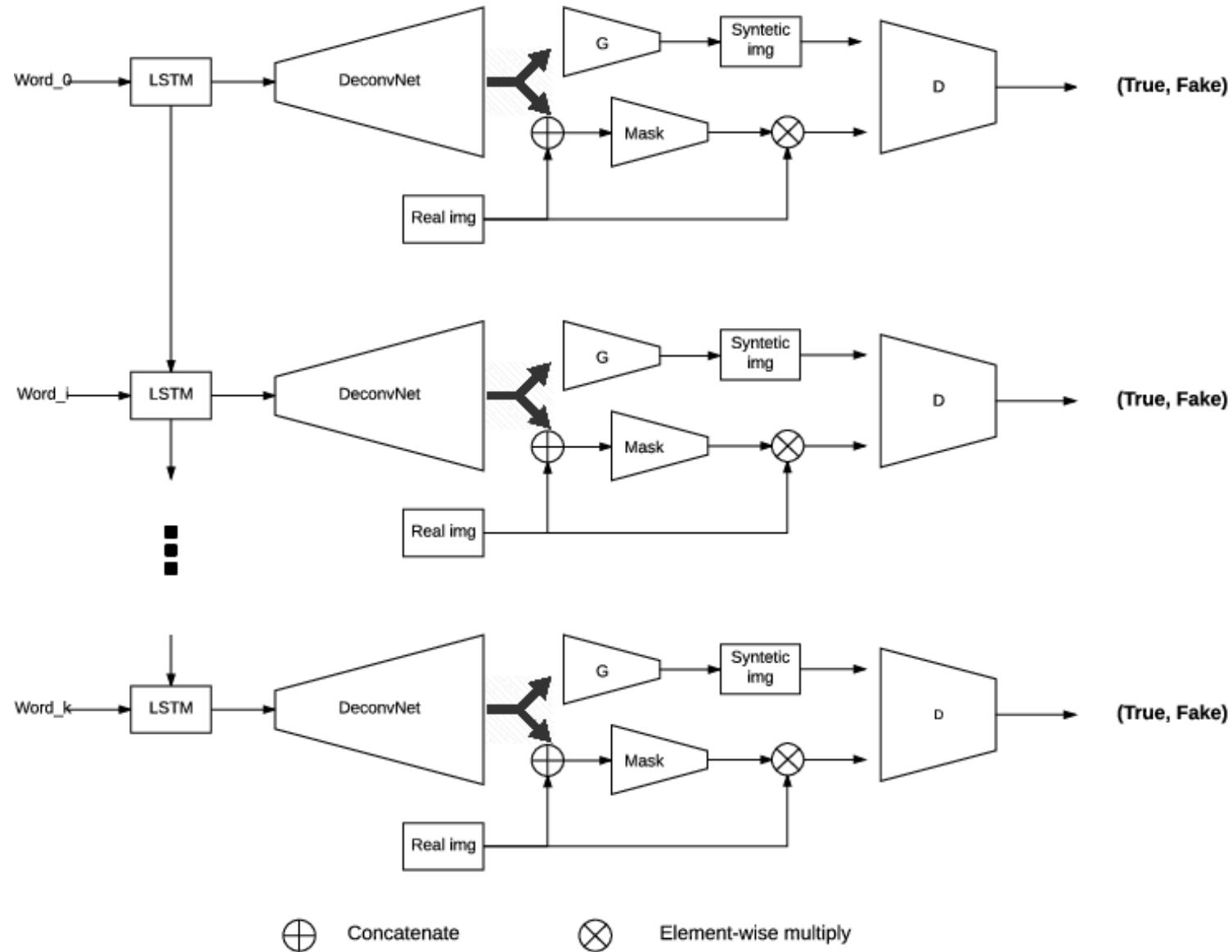
CGMOS is statistically significantly better than the compared methods regardless of classifiers selected.

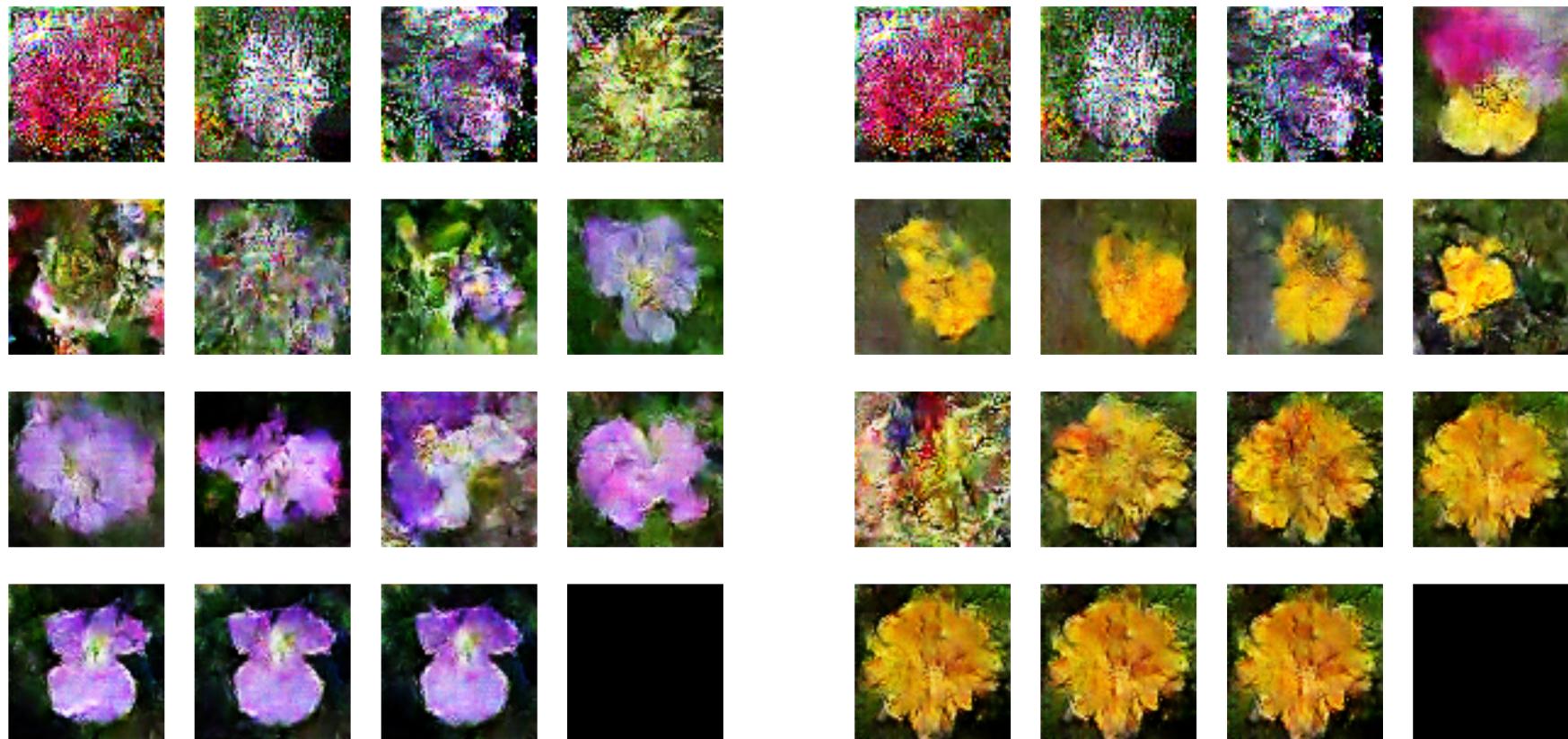
Text to image synthesis [22]

2.4.3.1 GAN. Generative adversarial networks are composed of a generator G and discriminator D , which are co-trained in a competitive manner. D learns to discriminate real images from fake ones generated by G , while G learns to generate fake images from latent variables z and tries to fool D . In other words, G and D play the following two-player minimax game with value function $V(G, D)$:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (2.2)$$

Text to image synthesis [22]





This flower has one large purple and white petal with one yellow and orange receptacle.

This flower has flat long and skinny yellow petals in one slightly upturned ring configuration.

