# Generative Correlation Discovery Network for Multi-Label Learning

**Lichen Wang, Zhengming Ding, Seungju Han, Jae-Joon Han, Changkyu Choi, Yun Fu**

wanglichenxj@gmail.com, zd2@iu.edu, {Sj75.han, jae-joon.han, changkyu_choi}@Samsung.com, yunfu@ece.neu.edu

SMILE Lab

Electrical & Computer Engineering

Northeastern University

# Multi-label Learning

**Background:**

- One object can be described by tens or hundreds of descriptions such as color, shape, texture and category. It is meaningful to predict all the labels from one instance.

**Goal:**

- Multi-label learning aims to seek a mapping from the feature space to the multi-label space.
  - Input: Instance samples (Image or features)
  - Output: multiple predicted labels


Input instance [1]

Label Space

Multi-label Classification Setting: Seeking a mapping from the feature space to the label space with multiple positive labels.

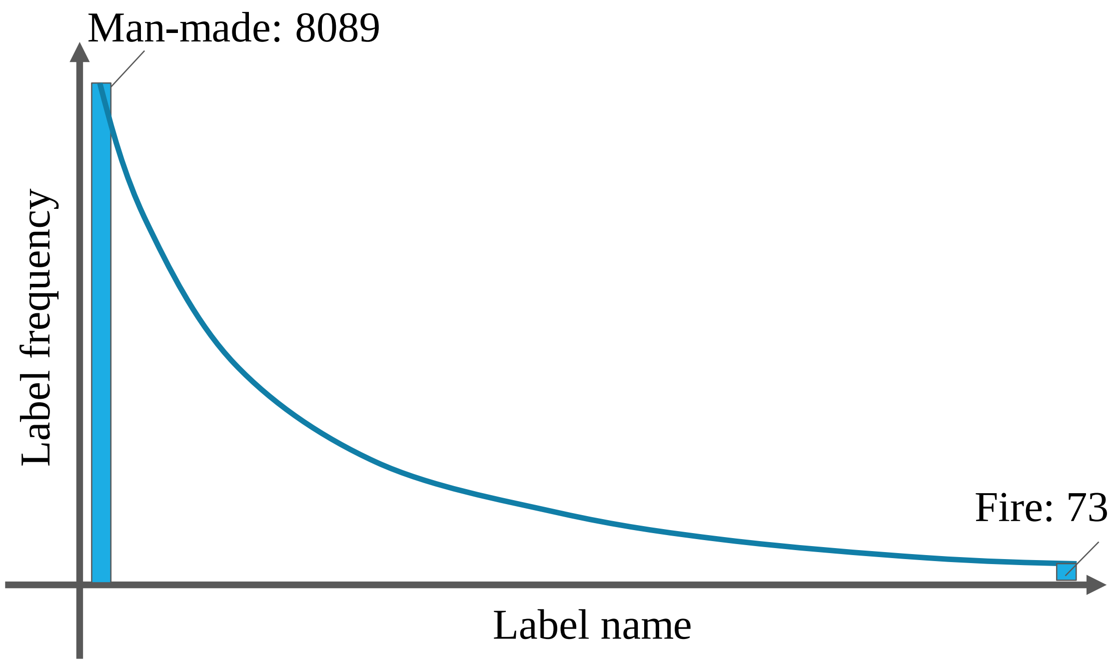
[1] Patterson, Genevieve, et al. "The SUN attribute database: Beyond categories for deeper scene understanding." International Journal of Computer Vision 108.1-2 (2014): 59-81.

# Challenges

## 1. Long-tail label distribution

The labels followed a long-tail distribution characteristic, which means some labels are always show up while some labels are rarely be assigned. The label imbalance significantly affect the learning performance in the multi-label learning setting.

E.g., For SUN dataset [1], the most common label (i.e., *Man-made*) shows up 8089 times in all the samples, while the rarest label (i.e., *Fire*) only shows up 73 times.



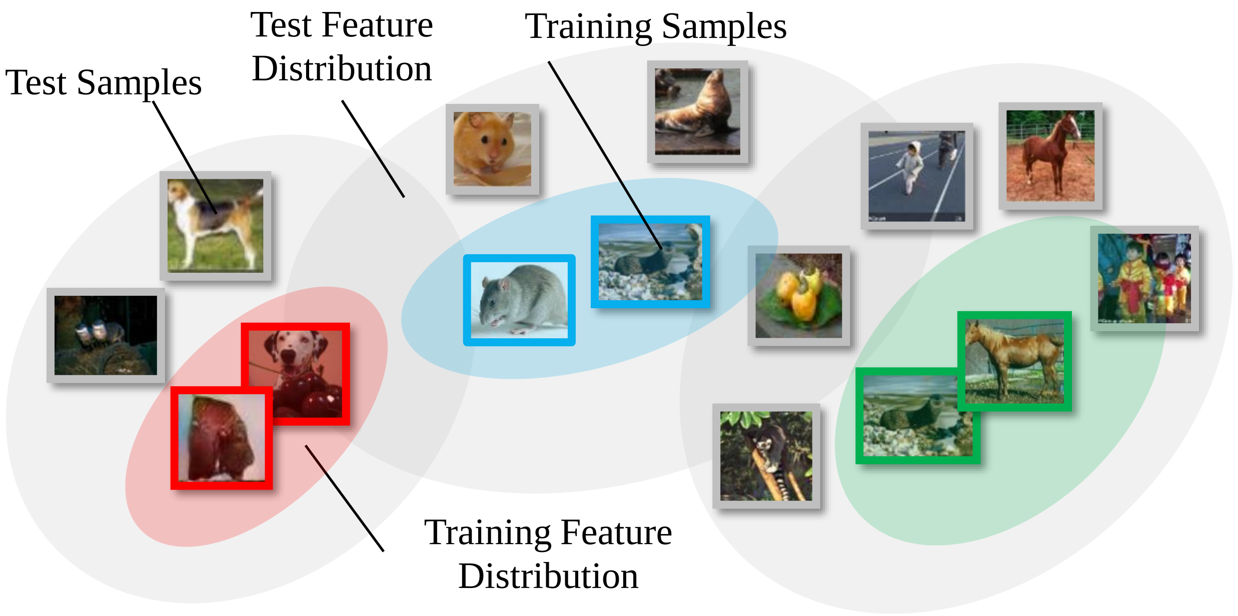

The curve of the label sample numbers. The difference is significant between the most common and rarest labels.

[1] Patterson, Genevieve, et al. "The SUN attribute database: Beyond categories for deeper scene understanding." IJCV108.1-2 (2014): 59-81

# Challenges

## 2. Distribution gaps between training and testing sets.

- The training sets cannot cover the entire feature space of testing set due to the small scale of the existing datasets and long-tail label distribution characteristic.
- This challenge is common in most machine learning tasks but more significant in multi-label learning scenario.
- How to efficiently utilize the label distribution is challenge for achieving high performance.

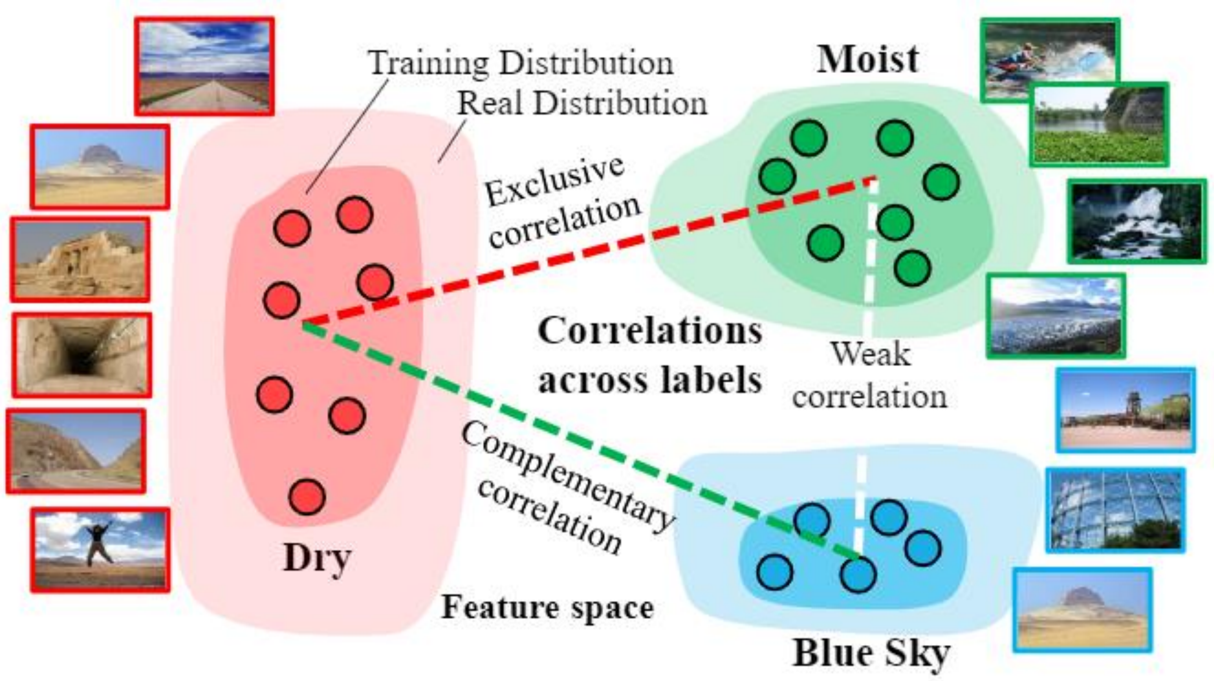



Distribution differences between training and testing features. Colored area denotes training set. Different colors indicate different labels. Gray area denotes testing set.

# Challenges

### 3. Complicated correlations across different labels:

- Label correlation is crucial for high accurate label prediction.
- Label correlation / semantic knowledge is difficult, expensive and challenging to obtain.
- Existing correlation is difficult to extend to a wide range of real-world applications.




Label correlation is crucial for accurate multi-label prediction. For instance, ***Dry*** and ***Moist*** are exclusive, while ***Dry*** and ***BlueSky*** are complementary.

# Model motivations

**Generative Correlation Discovery Network for Multi-Label Learning:**

We propose an end-to-end Generative Correlation Discovery Network (GCDN) method for multi-label learnin.

- GCDN captures the existing data distribution and synthesizes diverse data to enlarge the diversity of the training features.
- GCDN learns the label correlations based on a specifically-designed, simple but effective correlation discovery network. It automatically discover the label correlations and considerately improve the label prediction accuracy.

**Contribution:**

1. A specifically-designed multi-label conditional feature generative strategy is proposed. It synthesizes and diversifies the feature space to improve the model robustness and generalization.

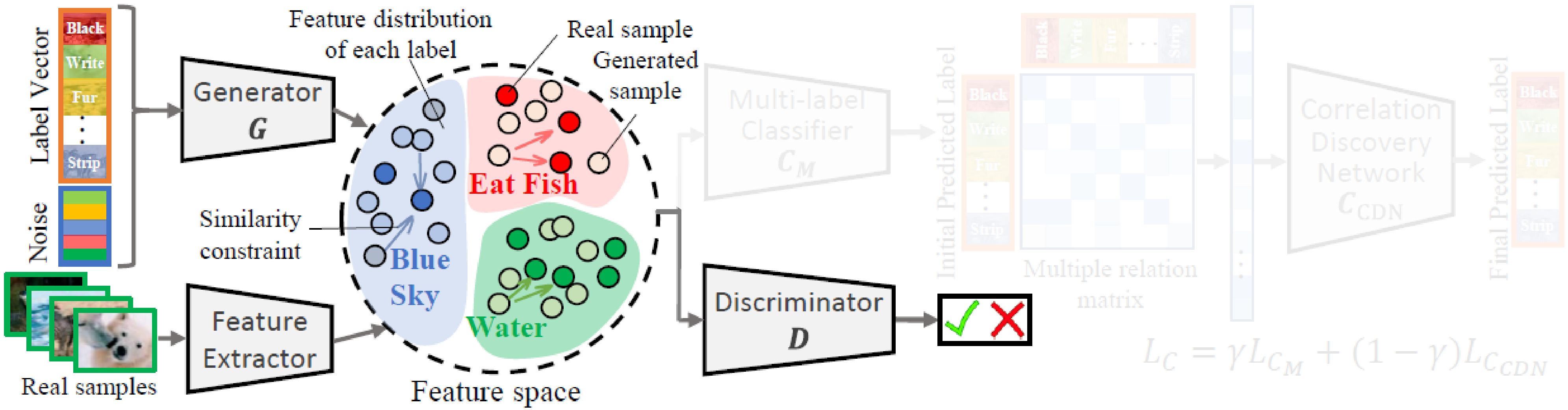
2. A graph-based Correlation Discovery Network (CDN) is proposed to automatically learn semantic correlations across different labels and utilize the knowledge to further improve learning performance.

# 1. Conditional Generative Module

- A conditional generative module is proposed. It generates new data conditioned on the input multi-label input.
- The conditioned label is extracted from ground truth in the generator training procedure
- A randomly generated label vector is utilized for training the following multi-label classifiers.

$$L_D = E_{X\sim p_X(X)} \log D(X|Y) + E_{z\sim p_z(z)} \log(1 - D(G(z|Y)))$$

$$L_{Gd} = -E_{z\sim p_z(z)} \log(1 - D(G(z|Y)))$$
$$L_{Gs} = \|G(z|Y) - X\|_F^2$$




Generative module to augment and diversify the training samples

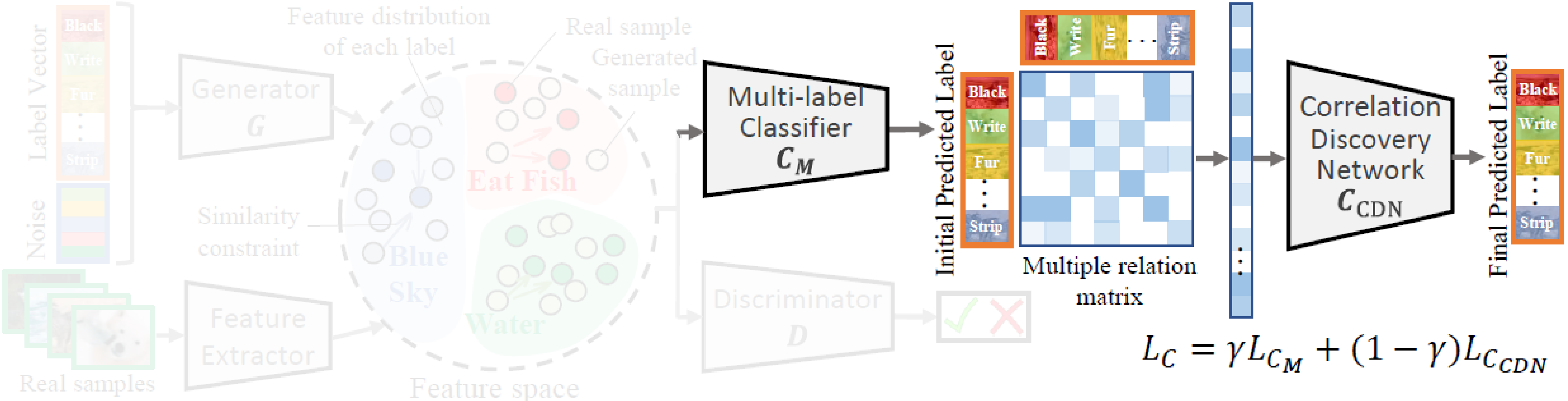
# 2. Correlation Discovery Network

- $C_M(.)$ obtains initial (low-accurate) results first, then $C_{CDN}(.)$ further utilizes the available prediction to "*tune*" the result to high-accurate..

$$L_{C_{CDN}} = \sum_{i=1}^{n_l} \|y_i - C_{CDN}(C_M(x_i)C_M(x_i)^\top)\|_2^2$$

- We balances the update processing between $C_M(.)$ and $C_{CDN}(.)$ to further help each other in the training stage and achieve the promising performance at last.

$$L_C = \gamma L_{C_M} + (1-\gamma)L_{C_{CDN}}$$



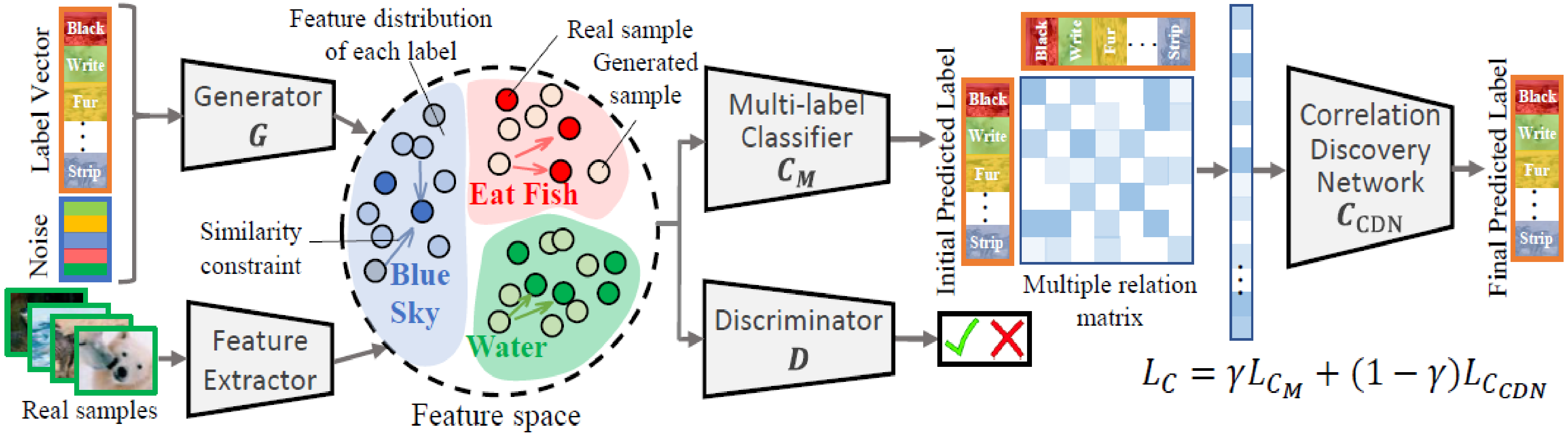

$$L_C = \gamma L_{C_M} + (1-\gamma)L_{C_{CDN}}$$

Generative module to augment and diversify the training samples

# Our approach

**Our complete model:**

- Generative model is utilized to generate and diversify the training samples. An adjacent matrix is built based on the predicted multi-label associated with another fully connected network.
- The model atomically learns the latent label correlation across different labels.
- All the networks are trained simultaneously to achieve the best performance.




Framework of our approach, where a generator, a discriminator and a classifier, and the correlation discovery network are simultaneously trained.

# Experiments

## Multi-label prediction performance:

- Six dataset are used for evaluate the multi-label recovery performance
  - Coreal, ESP, IAP, SUN, CUB, AWA
- Five evaluation metrics
  - Precision (Pre)
  - Recall (Rec)
  - F1
  - Non-zero Recall (N-R)
  - Mean average precision (mAP)

| Data | Method | Pre | Rec | F1 | N-R | mAP |
|---|---|---|---|---|---|---|
| Corel | LR | 0.2859 | 0.3211 | 0.3025 | 128 | 0.3630 |
| | SSMLDR | 0.2741 | 0.3366 | 0.3022 | 143 | 0.3410 |
| | FastTag | 0.3123 | 0.3657 | 0.3369 | 143 | 0.3871 |
| | ML-PGD | 0.2575 | 0.2911 | 0.2732 | 122 | 0.3727 |
| | SAE | 0.2962 | 0.3442 | 0.3184 | 141 | 0.3823 |
| | AG2E | 0.3011 | 0.3520 | 0.3245 | **157** | 0.3568 |
| | Ours | **0.3335** | **0.3714** | **0.3514** | 148 | **0.4417** |
| ESP | LR | 0.3793 | 0.2038 | 0.2653 | 215 | 0.3440 |
| | SSMLDR | 0.3298 | 0.1885 | 0.2399 | 226 | 0.3156 |
| | FastTag | 0.4011 | 0.1927 | 0.2617 | 208 | 0.3904 |
| | ML-PGD | 0.3239 | 0.2012 | 0.2482 | 210 | 0.4077 |
| | SAE | 0.3861 | 0.1743 | 0.2402 | 194 | 0.3842 |
| | AG2E | 0.3548 | 0.1525 | 0.2133 | 213 | 0.3730 |
| | Ours | **0.4032** | **0.2178** | **0.2828** | **239** | **0.4327** |
| IAP | LR | 0.4287 | 0.2041 | 0.2765 | 199 | 0.4211 |
| | SSMLDR | 0.3491 | 0.2520 | 0.2927 | 229 | 0.3981 |
| | FastTag | 0.4346 | 0.2267 | 0.2980 | 227 | 0.4596 |
| | ML-PGD | 0.4132 | 0.2441 | 0.3011 | 230 | 0.4674 |
| | SAE | 0.3537 | 0.2282 | 0.2774 | 213 | 0.4309 |
| | AG2E | 0.3829 | 0.2330 | 0.2897 | 229 | 0.4353 |
| | Ours | **0.4732** | **0.2648** | **0.3396** | **237** | **0.5295** |

| | | | | | | |
|---|---|---|---|---|---|---|
| SUN | LR | 0.6209 | 0.1473 | 0.2457 | 102 | 0.6807 |
| | SSMLDR | 0.6879 | 0.1700 | 0.2726 | 102 | 0.6723 |
| | FastTag | 0.6816 | 0.1473 | 0.2457 | 102 | 0.6914 |
| | ML-PGD | 0.7110 | 0.1614 | 0.2631 | 101 | 0.7087 |
| | SAE | 0.7183 | 0.1638 | 0.2668 | 98 | 0.7012 |
| | AG2E | 0.7685 | 0.1765 | 0.2871 | 99 | 0.6778 |
| | Ours | **0.7985** | **0.1835** | **0.2985** | **102** | **0.7093** |
| CUB | LR | 0.2010 | 0.0239 | 0.0428 | 157 | 0.0638 |
| | SSMLDR | 0.3410 | 0.0473 | 0.0832 | 178 | 0.2329 |
| | FastTag | 0.2147 | 0.0359 | 0.0615 | 167 | 0.3144 |
| | ML-PGD | 0.3334 | 0.0451 | 0.0794 | 155 | 0.3288 |
| | SAE | 0.3383 | 0.0514 | 0.0908 | 196 | 0.3255 |
| | AG2E | 0.3409 | 0.0531 | 0.0911 | 190 | 0.3106 |
| | Ours | **0.3718** | **0.0541** | **0.0944** | **214** | **0.3561** |
| AWA | LR | 0.8798 | 0.0821 | 0.1500 | 75 | 0.8626 |
| | SSMLDR | 0.7812 | 0.0858 | 0.1546 | 67 | 0.8346 |
| | FastTag | 0.7861 | 0.0949 | 0.1694 | 72 | 0.8791 |
| | ML-PGD | 0.5395 | 0.0635 | 0.1136 | 57 | 0.9121 |
| | SAE | 0.9683 | **0.0957** | **0.1742** | 73 | **0.9397** |
| | AG2E | 0.8483 | 0.0827 | 0.1507 | 73 | 0.9033 |
| | Ours | **0.9716** | 0.0871 | 0.1599 | **83** | 0.9291 |

Multi-label classification performance on six datasets

# Experiments

### Augmented datasets:

The work of [1] proposes a complete/augmented label set

- Corel: Averagely 3.40 to 4.84 labels
- ESP: Averagely 4.69 to 7.27 labels

### Zero-shot learning scenario:

The classes in training and testing sets are non-overlapped. It is a more challenging task since the distribution gaps are more significant.

| Data | Methods | Pre | Rec | F1 | N-R | mAP |
|---|---|---|---|---|---|---|
| Corel-A | LR | 0.2842 | 0.2304 | 0.2545 | 103 | 0.3762 |
| | SSMLDR | 0.3036 | 0.2791 | 0.2908 | 134 | 0.3660 |
| | FastTag | 0.3329 | 0.3145 | 0.3234 | 136 | 0.4127 |
| | ML-PGD | 0.3245 | 0.3011 | 0.3124 | 140 | 0.4275 |
| | SAE | 0.3168 | 0.3037 | 0.3101 | 128 | 0.4192 |
| | AG2E | 0.3273 | 0.3172 | 0.3221 | **143** | 0.3985 |
| | Ours | **0.3438** | **0.3219** | **0.3325** | 138 | **0.4773** |
| ESP-A | LR | 0.3848 | 0.1256 | 0.1894 | 178 | 0.3913 |
| | SSMLDR | 0.3253 | 0.1697 | 0.2231 | 202 | 0.3357 |
| | FastTag | 0.3886 | 0.1531 | 0.2197 | 196 | 0.4254 |
| | ML-PGD | 0.3713 | 0.1184 | 0.1795 | 162 | 0.4211 |
| | SAE | 0.3153 | 0.1425 | 0.1966 | 156 | 0.4050 |
| | AG2E | 0.3518 | 0.1492 | 0.2095 | 196 | 0.4030 |
| | Ours | **0.4772** | **0.1944** | **0.2763** | **225** | **0.4436** |

Multi-label Learning performance on label augmented Corel and ESP datasets

| Data | Method | Pre | Rec | F1 | N-R | mAP |
|---|---|---|---|---|---|---|
| SUN | LR | 0.7047 | 0.1548 | 0.2539 | 97 | 0.6616 |
| | SSMLDR | 0.6637 | 0.1481 | 0.2422 | 95 | 0.6581 |
| | FastTag | 0.6906 | 0.1522 | 0.2494 | 90 | 0.6706 |
| | ML-PGD | 0.7037 | 0.1471 | 0.2433 | 95 | 0.6829 |
| | SAE | 0.6978 | 0.1710 | 0.2747 | 100 | 0.6513 |
| | AG2E | 0.7125 | 0.1618 | 0.2637 | 88 | 0.6693 |
| | Ours | **0.7531** | **0.1857** | **0.2979** | **101** | **0.6911** |
| CUB | LR | 0.2600 | 0.0307 | 0.0549 | 160 | 0.2693 |
| | SSMLDR | 0.2926 | 0.0383 | 0.0677 | 166 | 0.2329 |
| | FastTag | 0.2231 | 0.0434 | 0.0726 | 143 | 0.2967 |
| | ML-PGD | 0.2392 | 0.0365 | 0.0635 | 117 | 0.3178 |
| | SAE | 0.2552 | 0.0469 | 0.0798 | 167 | 0.3102 |
| | AG2E | 0.2808 | 0.0481 | 0.0821 | 163 | 0.2693 |
| | Ours | **0.3091** | **0.0488** | **0.0843** | **179** | **0.3264** |
| AWA | LR | 0.7555 | 0.0766 | 0.1392 | 66 | 0.8809 |
| | SSMLDR | 0.7017 | 0.0764 | 0.1378 | 66 | 0.7858 |
| | FastTag | 0.8610 | 0.0912 | 0.1649 | 81 | 0.8918 |
| | ML-PGD | 0.4338 | 0.0623 | 0.1091 | 49 | 0.8677 |
| | SAE | 0.9015 | **0.0926** | **0.1679** | 78 | **0.8918** |
| | AG2E | 0.8247 | 0.0811 | 0.1476 | 71 | 0.8874 |
| | Ours | **0.9249** | 0.0804 | 0.1480 | **83** | 0.8784 |

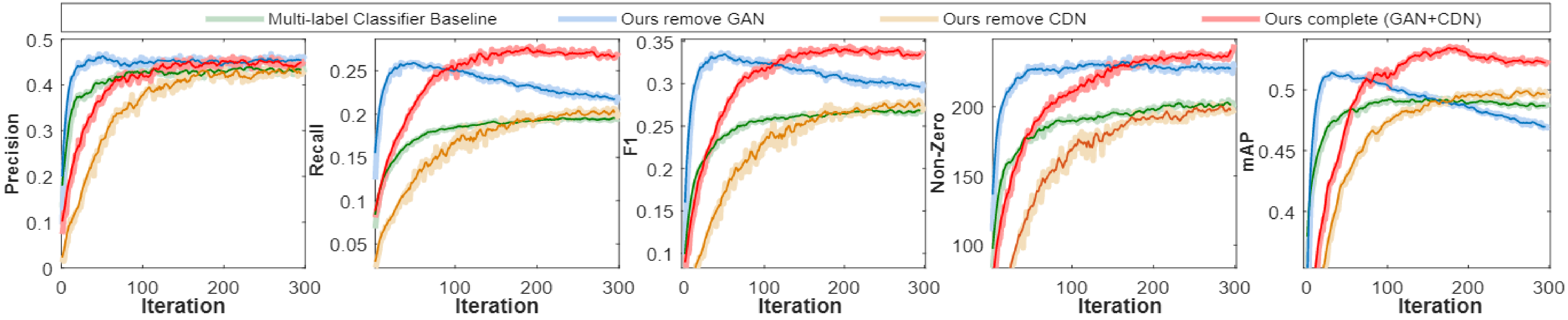
Multi-label Learning in zero-shot learning setting

[1] Wu, Baoyuan, Siwei Lyu, and Bernard Ghanem. "ML-MG: Multi-label learning with missing labels using a mixed graph.“ IEEE ICCV 2015

# Experiments

**Ablation study:**

- We run our approach based on 1) completed, 2) without GAN, 3) without CDN, and 4) without both GAN and CDN modules.
- The performance curves as the iteration increase are shown. We observe that the complete model achieves the highest performance. The CDN significantly improves the performance and the GAN model slightly improves the performance while significantly stabilizes the performance. We assume only CDN could easily cause overfitting due to the long-tail label distribution. Generative model effectively extend the feature distribution and improve the final prediction performance.

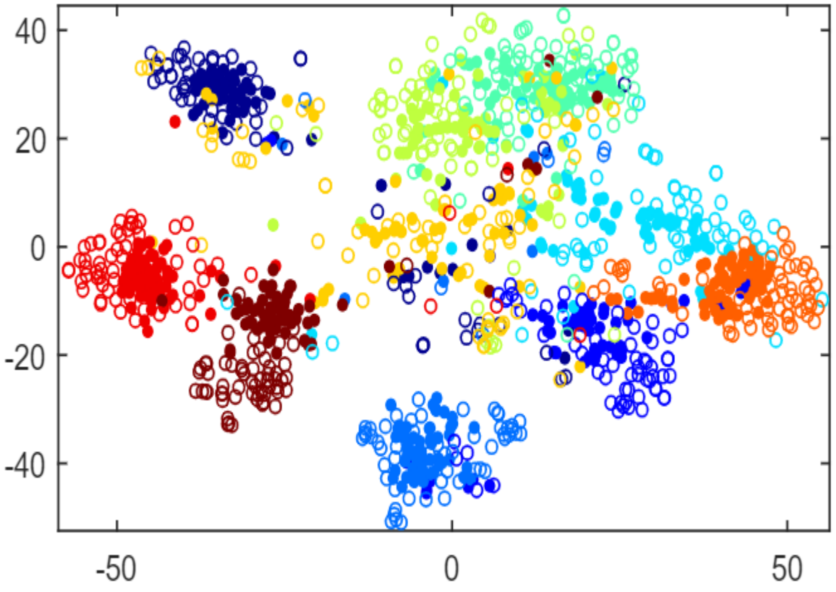



Ablation study of the proposed model. The curves indicate the performance when different modules are removed. It indicates the effectiveness of all the proposed modules in our GCDN model.

# Experiments

**Ablation study:**

- Visualization result of the real and generated samples in 2D space by t-SNE visualization approach. It shows that the real and generated samples are similar which helpful for sample augmentation and diversification.
- The experiment indicates that simply add noise in the feature space cannot improve the performance, which further demonstrates the effectiveness of generative strategy.




Real and generated samples in visual feature space

TABLE
MULTI-LABEL LEARNING PERFORMANCE OF ADDING VARIOUS LOW-LEVEL GAUSSIAN NOISES TO THE ORIGINAL FEATURE OF CUB DATASET.

| Noise | Pre | Rec | F-1 | N-R | mAP |
|---|---|---|---|---|---|
| 0.00 | **0.3718** | **0.0541** | **0.0944** | **214** | **0.3561** |
| 0.05 | 0.3711 | 0.0540 | 0.0941 | 214 | 0.3561 |
| 0.10 | 0.3692 | 0.0538 | 0.0943 | 214 | 0.3537 |
| 0.15 | 0.3668 | 0.0537 | 0.0941 | 214 | 0.3511 |
| 0.20 | 0.3647 | 0.0534 | 0.0938 | 212 | 0.3482 |
| 0.25 | 0.3612 | 0.0533 | 0.0936 | 211 | 0.3467 |
| 0.30 | 0.3591 | 0.0531 | 0.0932 | 209 | 0.3416 |
| 0.35 | 0.3505 | 0.0530 | 0.0930 | 208 | 0.3389 |
| 0.40 | 0.3393 | 0.0529 | 0.0929 | 206 | 0.3351 |
| 0.45 | 0.3314 | 0.0528 | 0.0927 | 204 | 0.3232 |
| 0.50 | 0.3248 | 0.0526 | 0.0926 | 202 | 0.3215 |

Multi-label performance when different level of Gaussian noise is added into the visual feature

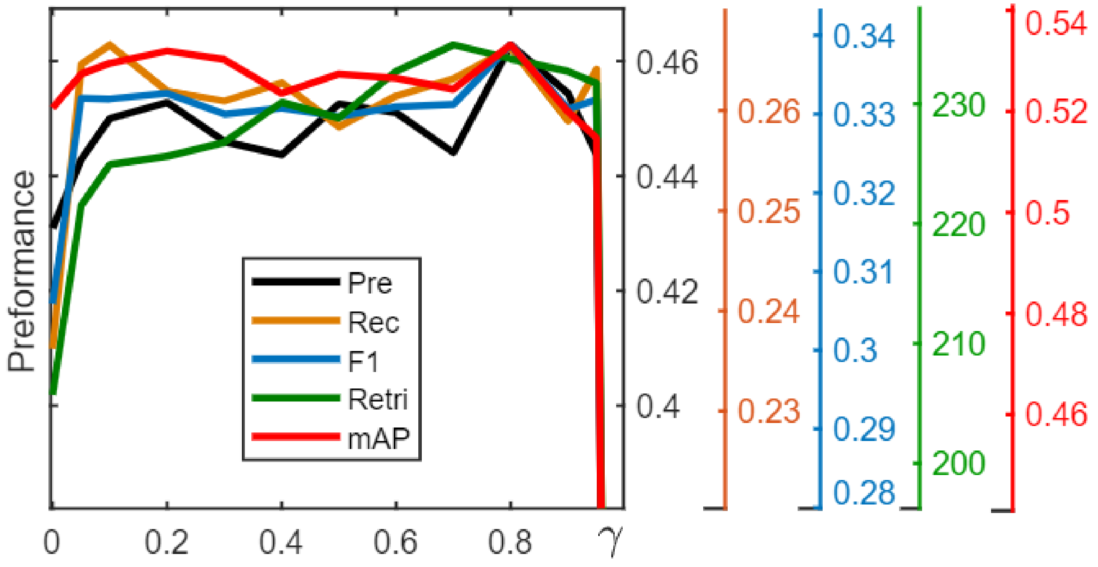
# Experiments

### Parameter sensitivity of $\gamma$:

Parameter $\gamma$ balances the weights between the multi-label classifier and the correlation discovery network:

- $\gamma = 1$, only multi-label classifier
- $\gamma = 0$, does not consider label correlations

$$L_C = \gamma L_{C_M} + (1-\gamma) L_{C_{CDN}}$$




Parameter sensitivity of $\gamma$

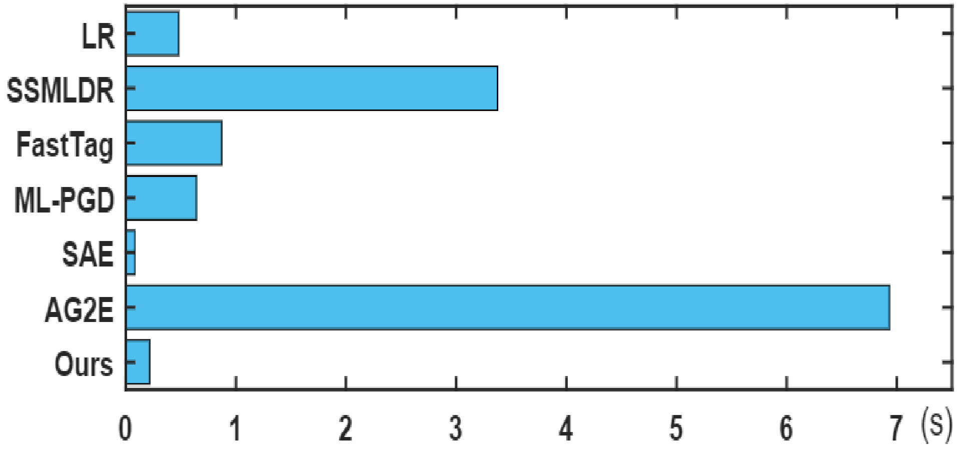
Conclusions:

- The performance achieves highest performance when $\gamma \in [0,1]$
- The proposed label correlation discovery network is effective for further improve the learning performance

# Experiments

**Time Consumption:**

- The approaches which require optimization procedure are computational costly (i.e., SSMLDR, AG2E).
- Other approach achieves relatively linear complicity which are fast (i.e., LR, FastTag, SAE)
- Although multiple layers are deployed in our model, the approach achieves the high speed due to the simple network structure and the GPU acceleration.

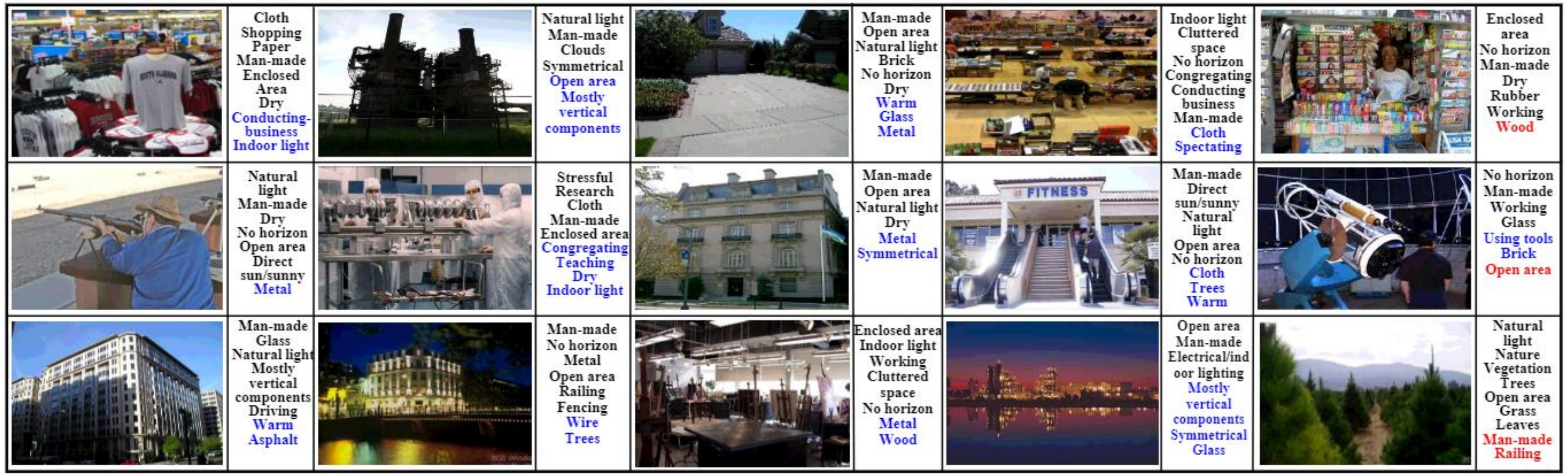



Time consumption in testing stage

# Image annotation

**Image annotation result:**

- The case study of the image annotation result. The images are extracted from SUN dataset [1]
- The results illustrates that our approach achieves high performance and even recovery the “missing” labels in the ground truth data.




Samples of recovered labels from SUN dataset. **Black** font denotes labels that match with the ground truth. **Blue** font denotes labels that do not exist in the ground truth but match our judgments. **Red** font denotes incorrect labels from our model. The result shows that our approach is robust and able to recover labels even when labels are missed from the ground truth.


[1] Patterson, Genevieve, et al. "The SUN attribute database: Beyond categories for deeper scene understanding." International Journal of Computer Vision 108.1-2 (2014): 59-81.

# Image retrieval

**Image retrieval results:**

- Given a target label (e.g., *Wood*), image retrieval extract all the relevant images from the dataset.
- The results shows high retrieval performance. The most of the top retrieval results are correct.

Phenomena:

- Adjective and verb labels are more challenging than noun labels.
- Model prefers specific scenes than others


Image retrieval result of SUN dataset [1] in zero-shot scenario. Each row shows the images with the highest corresponding label score retrieved from the test set. **Green** and **red** boxes indicate correct and incorrect retrieval, respectively. For each target label, we show incorrect retrieval result and its score rankings on the image right corner.

[1] Patterson, Genevieve, et al. "The SUN attribute database: Beyond categories for deeper scene understanding." International Journal of Computer Vision 108.1-2 (2014): 59-81.

# GCDN Summary

**Project goal**

- **Generative Correlation Discovery Network for multi-label learning (GCDN):**
  - Learn feature distribution & Generative samples, fill up distribution gap.
  - Automatically explore semantic label correlations and further improve the learning performance.
- **Four experimental settings:**
  - Multi-label classification
  - Zero-shot Multi-label classification
  - Automatic image annotation
  - Image retrieval
- **Six image datasets:**
  - SUN, AWA, CUB, Core5K, ESP Game, IAPRTC-12
- **Experimental results:**
  - Quantified & visualization result
  - Ablation study demonstrates the effectiveness of the model
  - Efficient in the testing process

# Questions?

Please contact: **wanglichenxj@gmail.com** for more questions.

# Thank you!

SMILE Lab
Electrical & Computer Engineering
Northeastern University