Development of Unsupervised-Learning Based Glitch Classification System

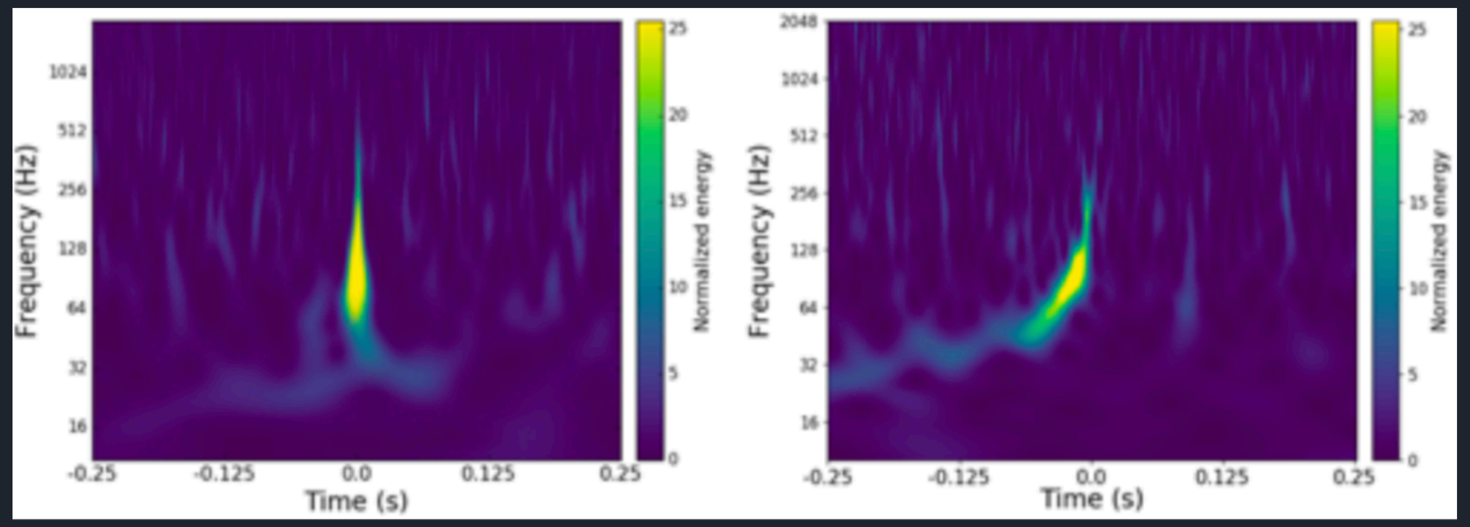
Gen Ueshima[1] ([1]Nagaoka University of Technology)

Keiko Kokeyama[2], Chihiro Kozakai[3], Shoichi Oshino[2], Yutaka Shikano[4], Hirotaka Takahashi[5], Takashi Uchiyama[2], Tatsuki Washimi[3], Takahiro Yamamoto[2], Takaaki Yokozawa[2]
([2]ICRR, [3]NAOJ, [4]Keio U, [5]TCU)

KAGRA International Workshop (KIW)

Background

- Gravitational wave telescopes have a large variety of glitch sources **<u>with corresponding time-frequency characteristics</u>**.



Different types of glitch, Blip (left) and Chirp (left)

- It is important to **classify** the time-frequency characteristics around glitches.
  - **We can get <u>statistical information</u> for each source of glitch to take some measures.**

- Applying **Supervised Learning** based on deep learning [1] to glitch have some aspects of effective, or issues (#1, #2)

<u>#1 Issues in general (glitch classification)</u>

- We don’t know the true number of glitch sources.
- Glitch has different frequencies (occur) depending on its source (Imbalanced data).

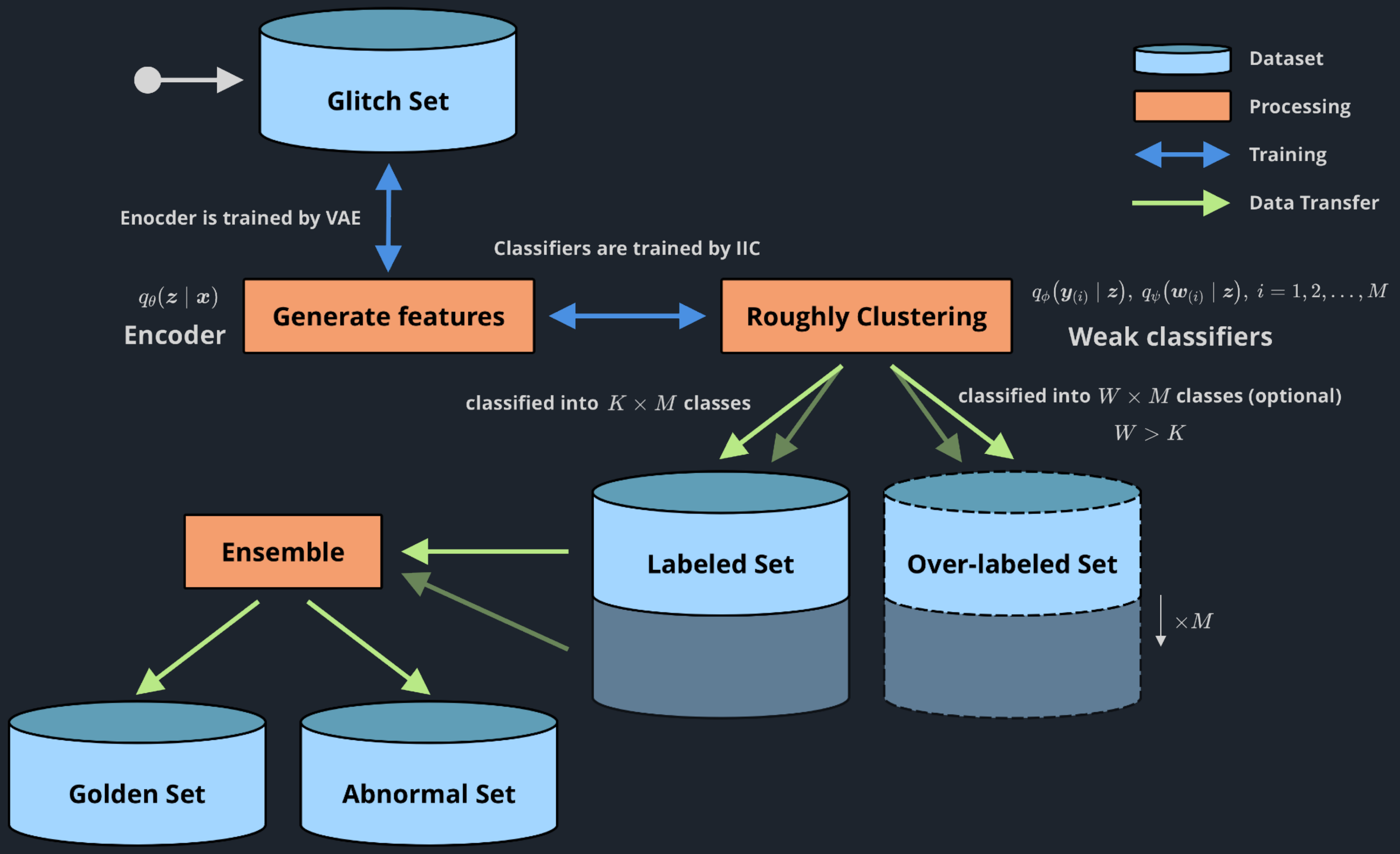
<u>#2 Issues depending on Supervised Learning</u>

- Highly dependent on skill of human to labeling. (requires a lot of work for labeling)
- Subclasses or abnormals may be included.

We need some human-independent system for classification overcoming #1, #2

Overview of our system

- We develop a system to classify glitches based on **Unsupervised Learning.**
- We tested the system by applying to ***Gravity Spy Dataset*** [1].



Architecture of our system

Classification strategy

1. Generate features from glitch set using **Variational Autoencoder (VAE)** [2]
2. Roughly classify features into arbitrary number of classes (K, W) by multiple weak classifiers $(\times M)$ using **Invariant Information Clustering (IIC)** [3]
   - **Labeled Set**：Using K classes, M patterns
   - **Over-labeled Set (optional)**：Using $W\ (> K)$ classes, M patterns
3. Ensemble (Consensus) multiple clustering results (preliminary)
   - **Golden Set**：Reliable glitches with reliable labels
   - **Abnormal Set**：Unreliable glitches

[1] M. Zevin, et al., "Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science", Classical and Quantum Gravity, Vol.34, 064003 (2017).
[2] D. Kingma, et.al, "Auto-Encoding Variational Bayes",arXiv:1312.6114v10 (2014).
[3] Xu Ji, et.al, "Invariant Information Clustering for Unsupervised Image Classification and Segmentation", arXiv:1807.06653v4 (2019).

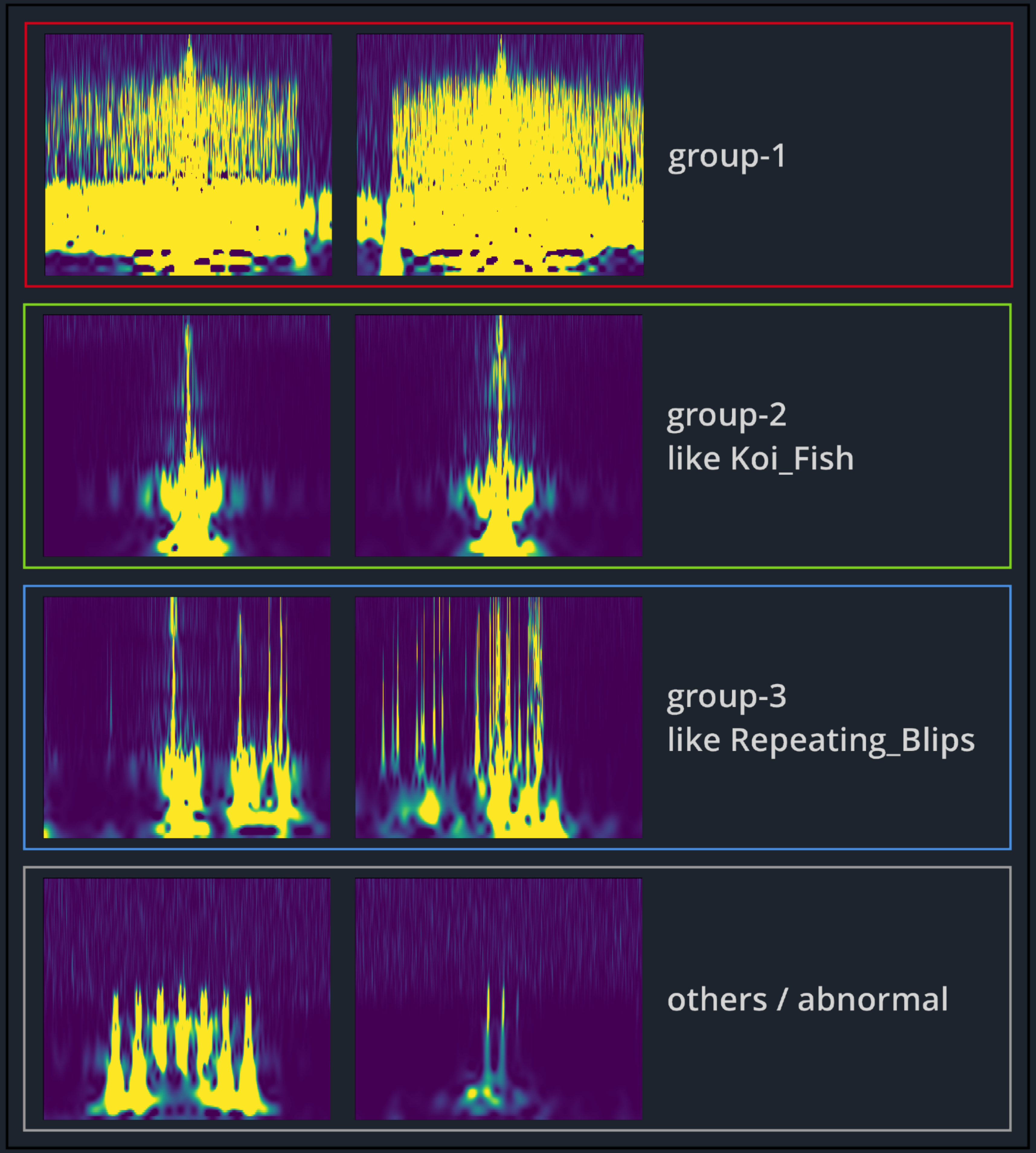
Consideration from *Gravity Spy Dataset*

- ***Gravity Spy Dataset*** have 22 types of glitches labeled by unique pipeline (human with CNN) [1].

<u>Consideration</u>

- Abnormals may be included
- Certain classes may be divided into multiple subclasses
- Some classes have too much diversity

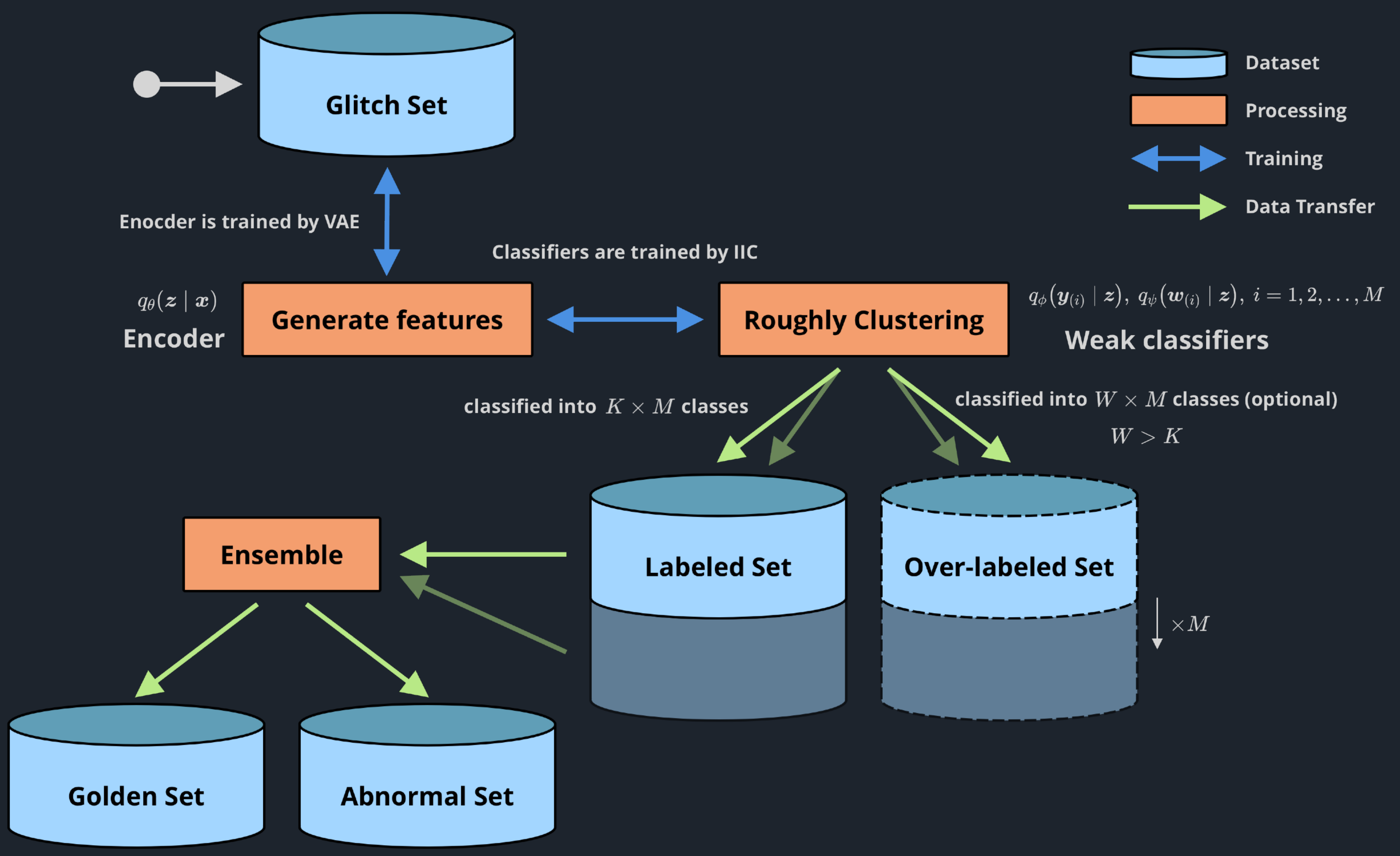
We designed the system to overcome these issues !



Glitches in Extremely_Loud class can divide into subclasses

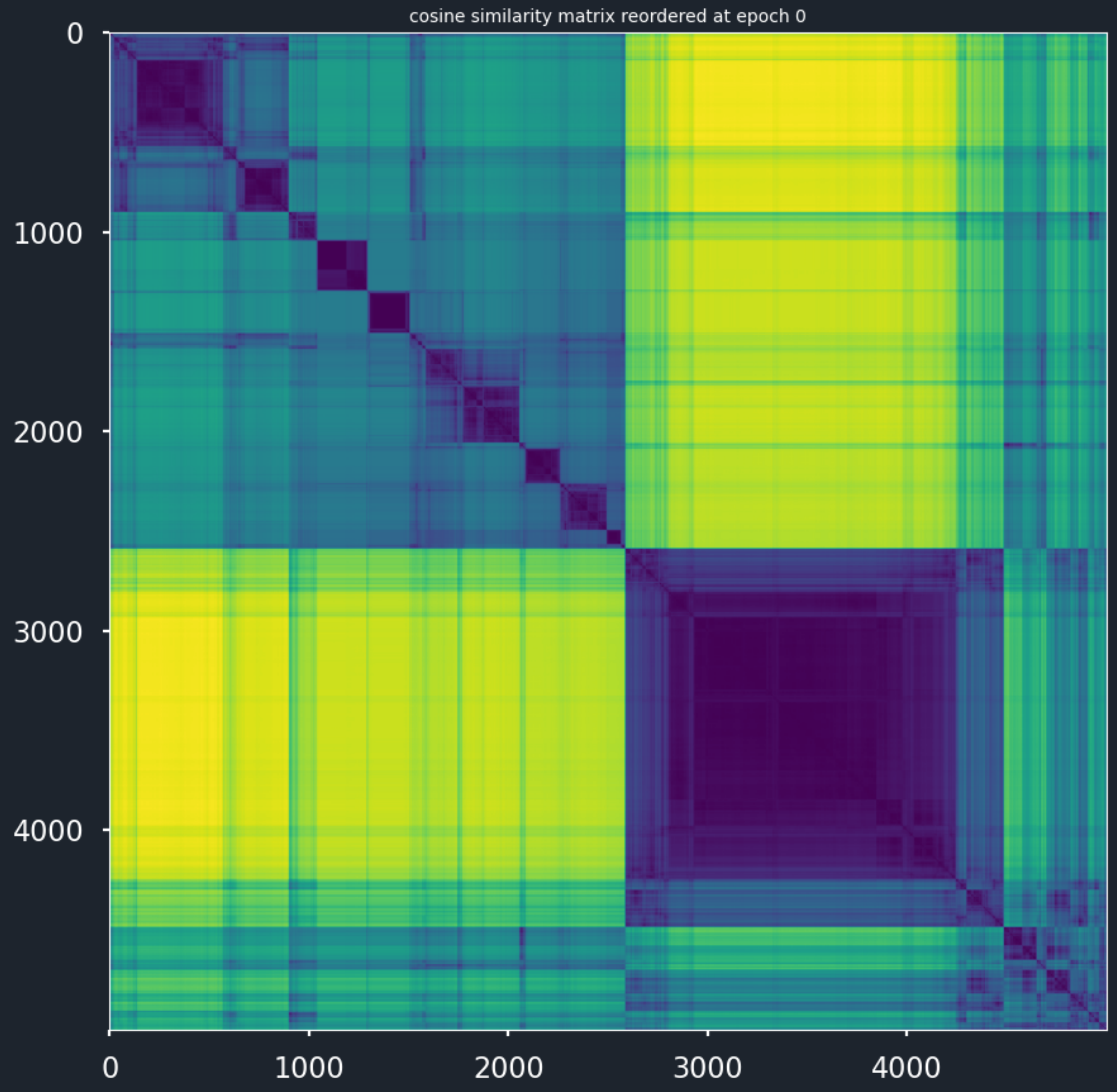
Overview of our system

- We develop a system to classify glitches based on **Unsupervised Learning.**

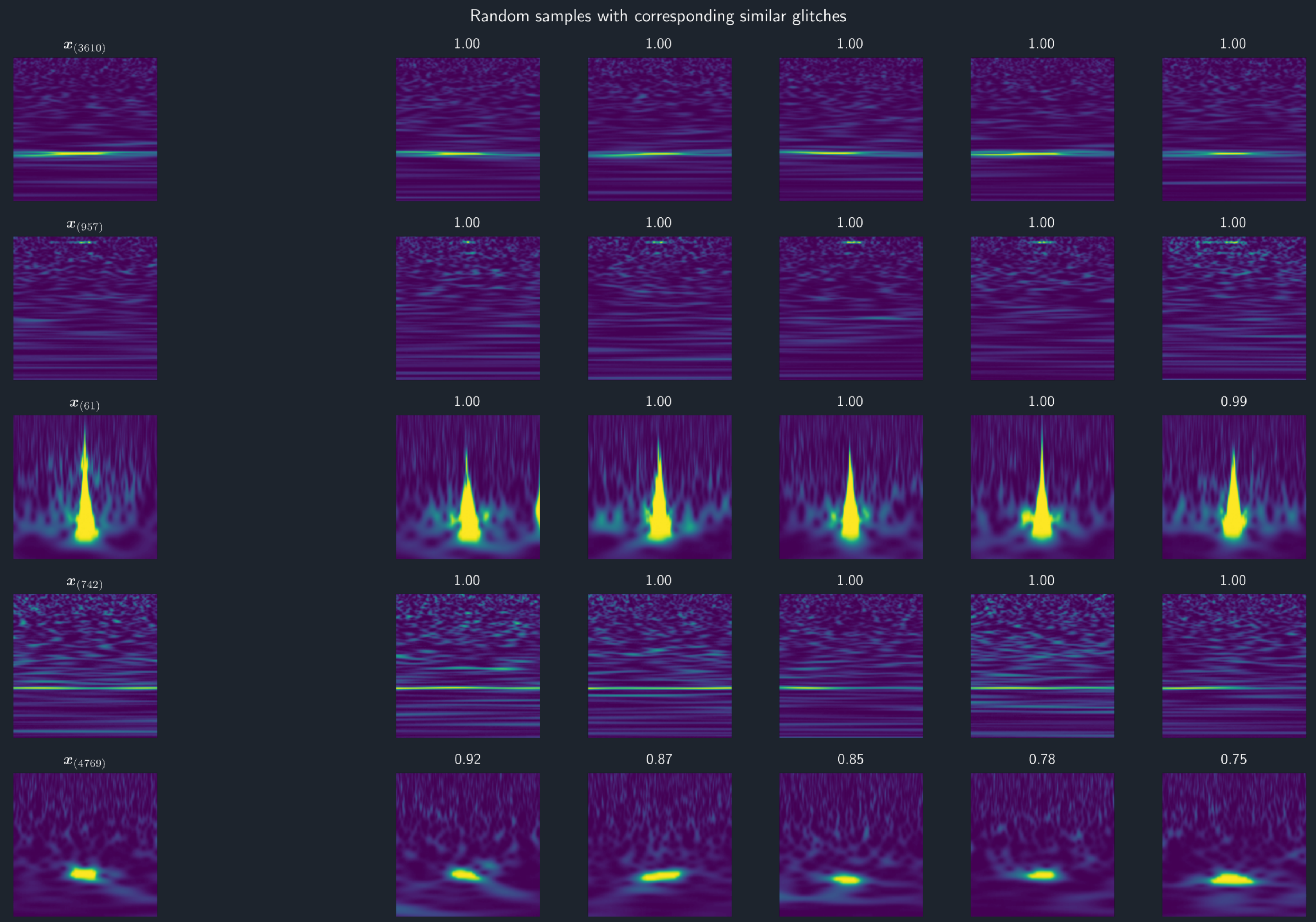


Architecture of our system

Overview of results



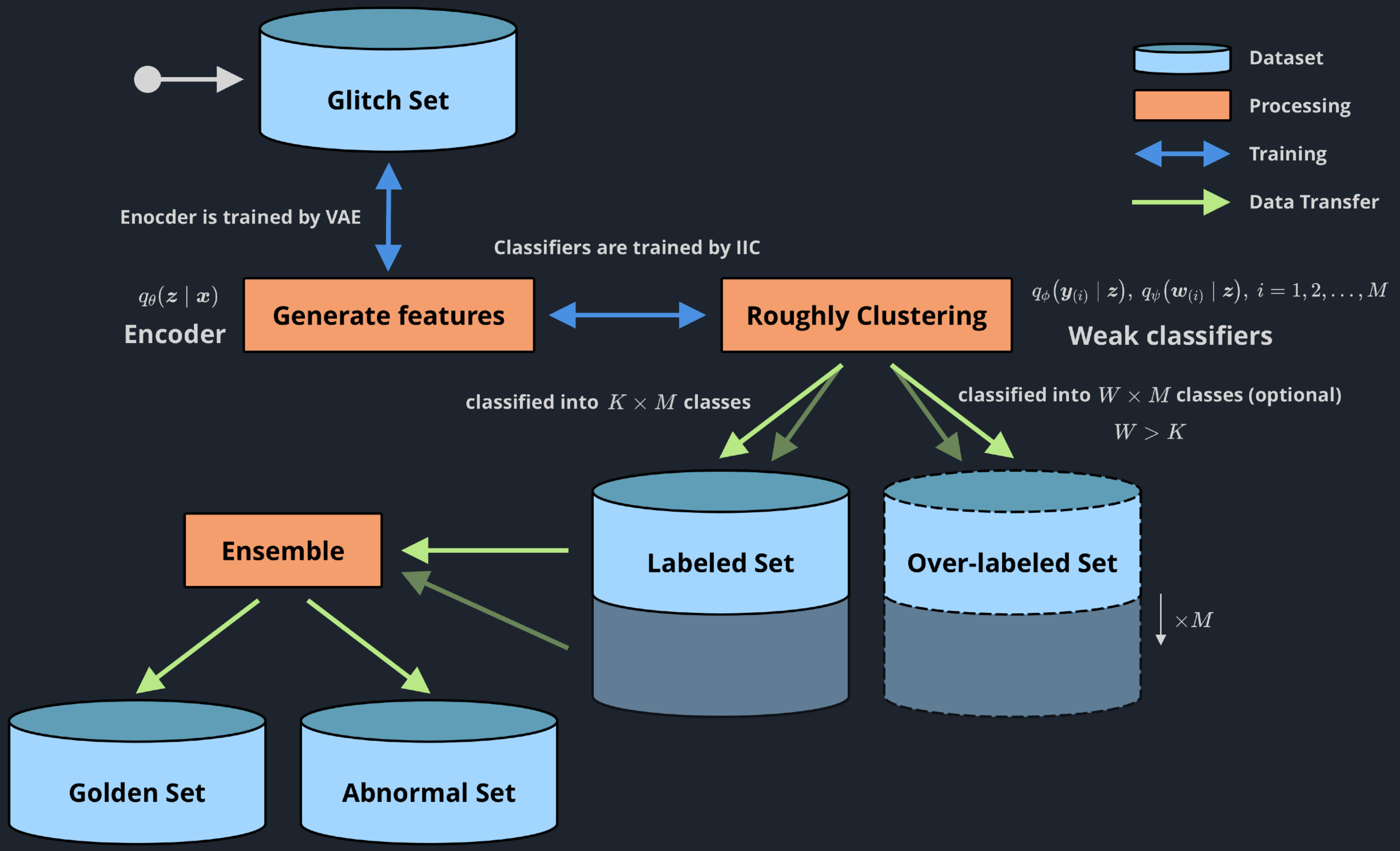
Sorted cosine distance matrix of hyper graph (test data)



Random samples (left) with the 5 most similar samples (right)

Overview of our system

- We develop a system (Fig. 1) to classify glitches based on **Unsupervised Learning.**



Architecture of our system

Classification strategy

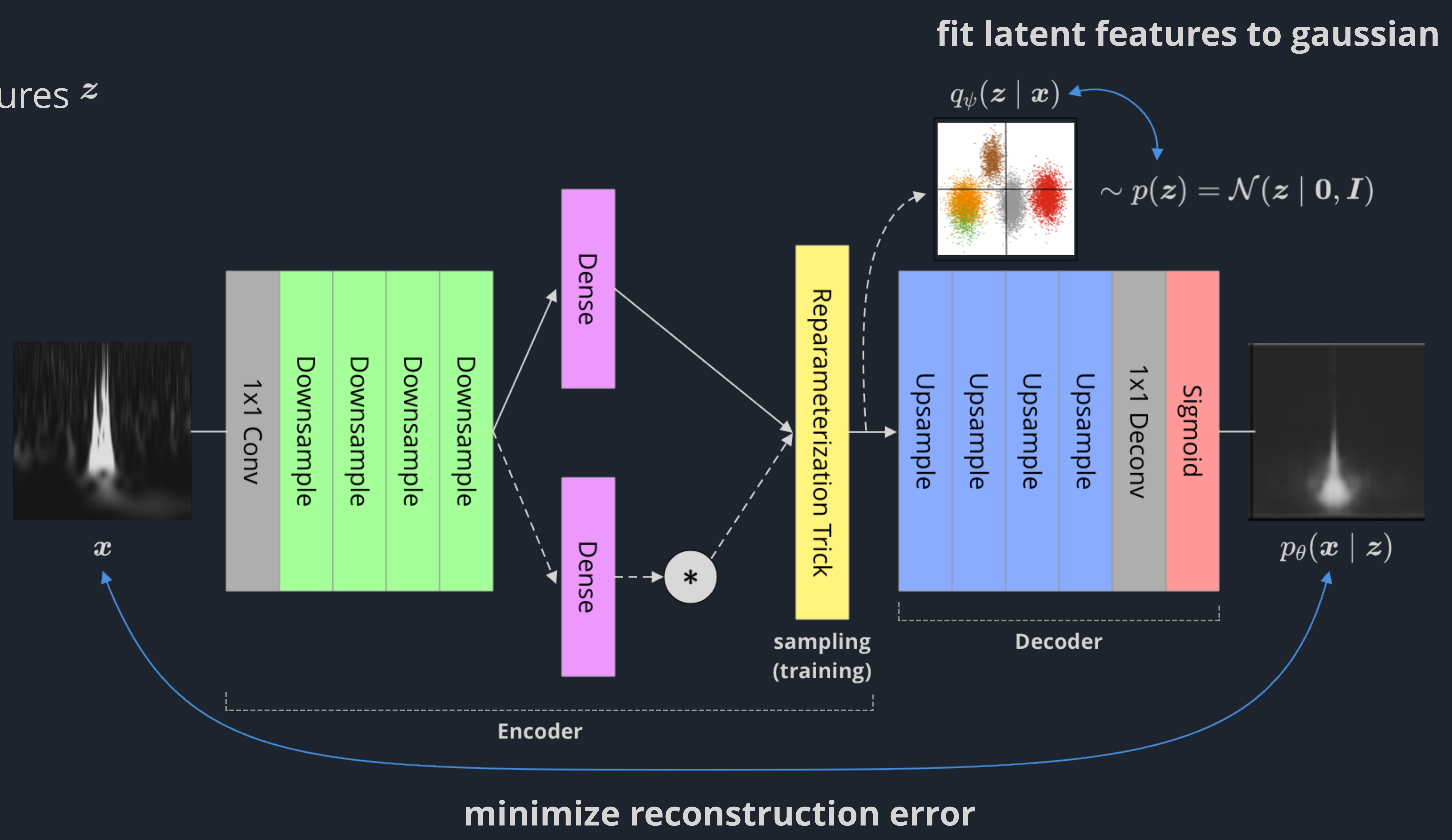
1. Generate features from glitch set using **Variational Autoencoder (VAE)** [2]
2. Roughly classify features into arbitrary number of classes (K, W) by multiple weak classifiers ($\times M$) using **Invariant Information Clustering (IIC)** [3]
   - **Labeled Set**：Using K classes, M patterns
   - **Over-labeled Set (optional)**：Using W ($> K$) classes, M patterns
3. Ensemble (Consensus) multiple clustering results (preliminary)
   - **Golden Set**：Reliable glitches and with hyper graph with reliable labels
   - **Abnormal Set**：Unreliable glitches

Variational Autoencoder (VAE) [2]

- Unsupervised or semi-supervised generative model by assuming, input data $\boldsymbol{x}$ is generated by latent features $\boldsymbol{z}$ (our system use unsupervised model)
- We can generate disentangled features $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ from input data $\boldsymbol{x}$

Architecture

- **Encoder** $q_\psi(\boldsymbol{z} \mid \boldsymbol{x})$ infer latent features $\boldsymbol{z}$ from input data $\boldsymbol{x}$
- **Decoder** $p_\theta(\boldsymbol{x} \mid \boldsymbol{z})$ generate reconstructed image $\boldsymbol{x}$ from latent features $\boldsymbol{z}$

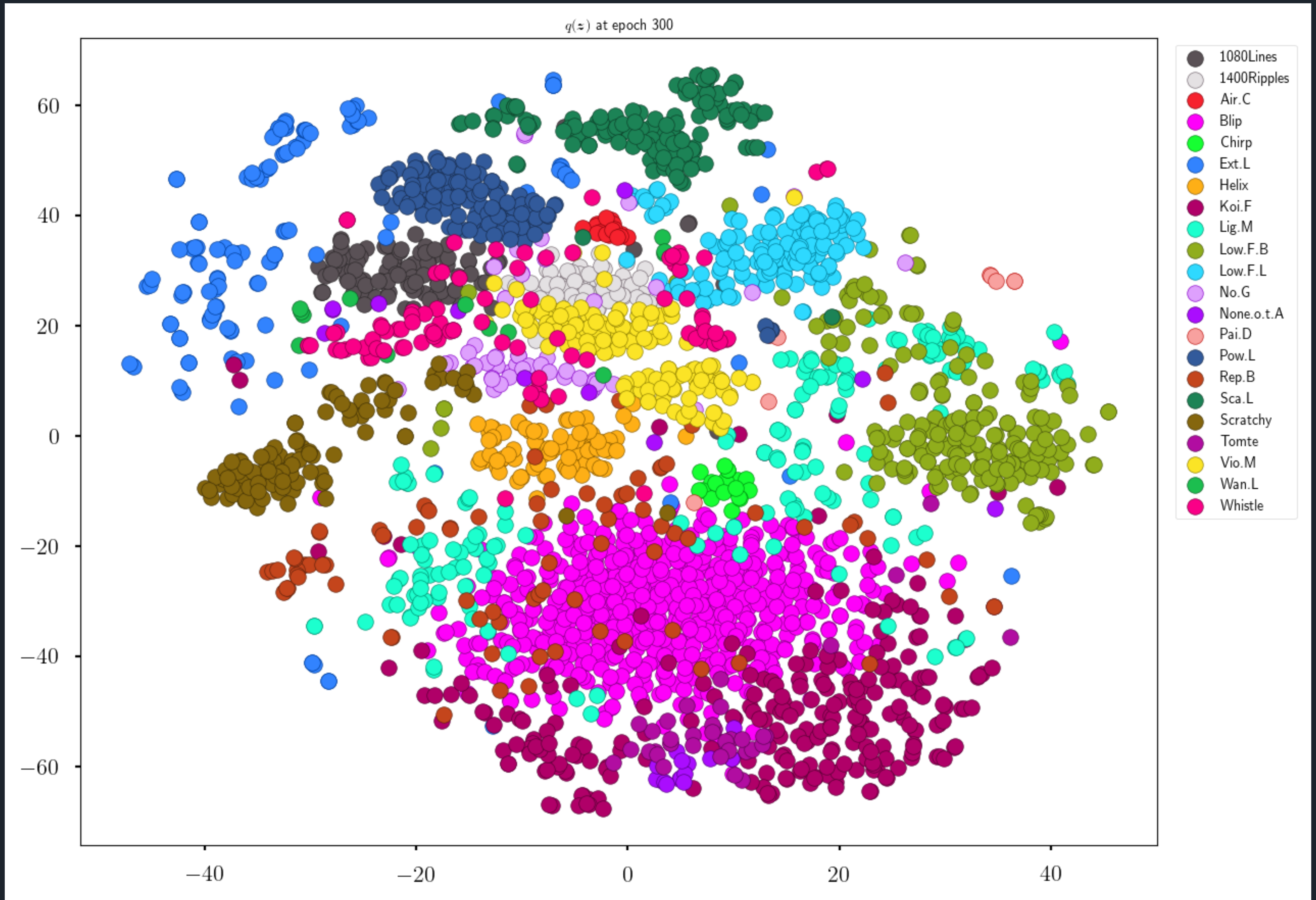


Objective

$$
\begin{aligned}
\text{maximize} \log p_\theta(\boldsymbol{x}) &= \log \mathbb{E}_{q_\psi(\boldsymbol{z}|\boldsymbol{x})}\left[\frac{p(\boldsymbol{z})p_\theta(\boldsymbol{x} \mid \boldsymbol{z})}{q_\psi(\boldsymbol{z} \mid \boldsymbol{x})}\right] \\
&\geq \mathbb{E}_{q_\psi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p(\boldsymbol{z})p_\theta(\boldsymbol{x} \mid \boldsymbol{z})}{q_\psi(\boldsymbol{z} \mid \boldsymbol{x})}\right] \\
&= \mathbb{E}_{q_\psi(\boldsymbol{z}|\boldsymbol{x})}[\log p_\theta(\boldsymbol{x} \mid \boldsymbol{z})] - \mathbb{E}_{q_\psi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{q_\psi(\boldsymbol{z} \mid \boldsymbol{x})}{p(\boldsymbol{z})}\right] \\
&= \frac{1}{L}\sum_{i=1}^{L} \log p_\theta(\boldsymbol{x} \mid \boldsymbol{z}_{(i)}) - \mathcal{D}[q_\psi(\boldsymbol{z} \mid \boldsymbol{x}) \| p(\boldsymbol{z})] = \mathcal{L}
\end{aligned}
$$

[2] D. Kingma, et.al, "Auto-Encoding Variational Bayes",arXiv:1312.6114v10 (2014).

Result-1: Latent features generated by VAE



2D latent features by t-SNE with true labels (Gravity Spy)

Consideration

- We identified the presence of classes containing multiple subclasses such as Ext.L, Vio.M.
- The data distribution for each class partly overlaps with the others, but it is generally separated.
- We found that differences in the statistical variance between the classes by visualizing the distribution of the data,

Invariant Information Clustering (IIC) [3]

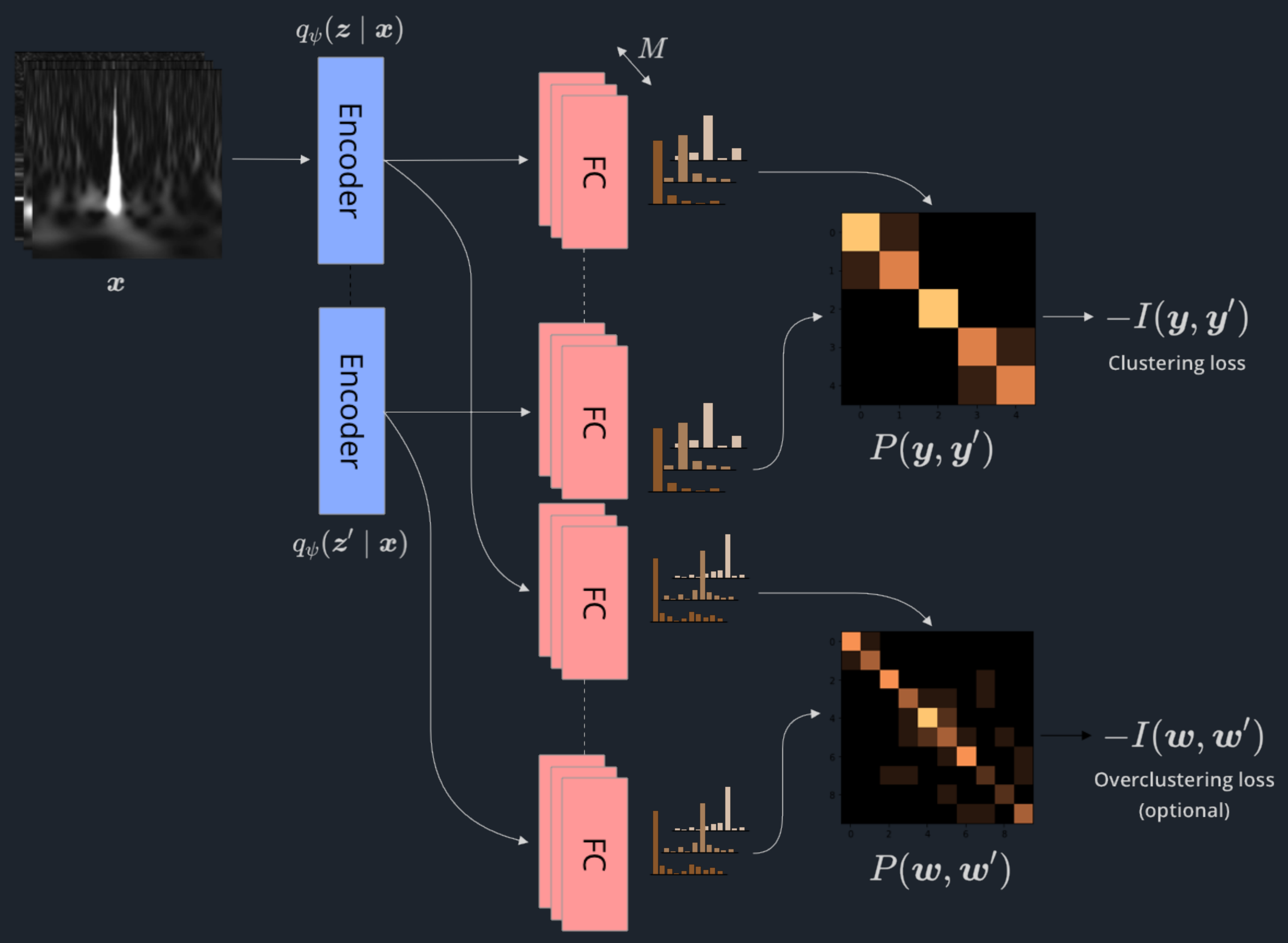
- Unsupervised classifier by maximize mutual information of classification results from two different types of features z, z'
- We can roughly classify features to arbitrary number of classes K or / and $W > K$, and can get different kinds M of results.

Architecture

- **Encoder** $q_\psi(\boldsymbol{z} \mid \boldsymbol{x})$ infer latent features $\boldsymbol{z}$ from input data $\boldsymbol{x}$
- **Multiple weak classifiers** classify $\boldsymbol{z}$ to arbitrary number of classes with different kinds of results.

Objective

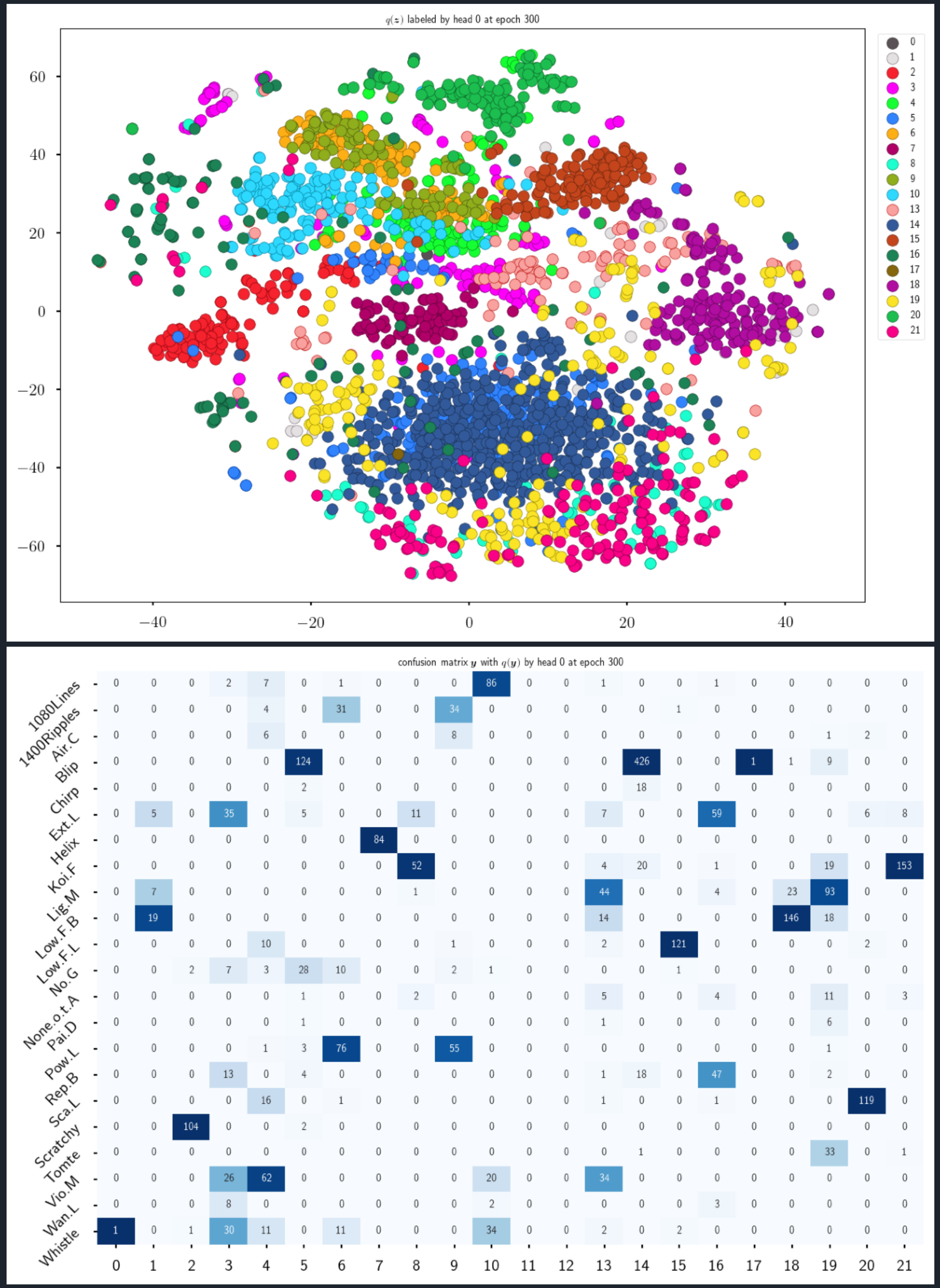
$$\max[I(\boldsymbol{y}, \boldsymbol{y}') + I(\boldsymbol{w}, \boldsymbol{w}')]$$
$$I(\boldsymbol{y}, \boldsymbol{y}') = \sum_{k=1}^{K} \sum_{k'=1}^{K} P(y = k, y' = k') \log \frac{P(y=k, y'=k')}{P(y=k|\boldsymbol{x}) P(y'=k'|\boldsymbol{x})}$$
$$I(\boldsymbol{w}, \boldsymbol{w}') = \sum_{c=1}^{C} \sum_{c'=1}^{C} P(w = c, w' = c') \log \frac{P(w=c, w'=c')}{P(w=c|\boldsymbol{x}) P(w'=c'|\boldsymbol{x})}$$



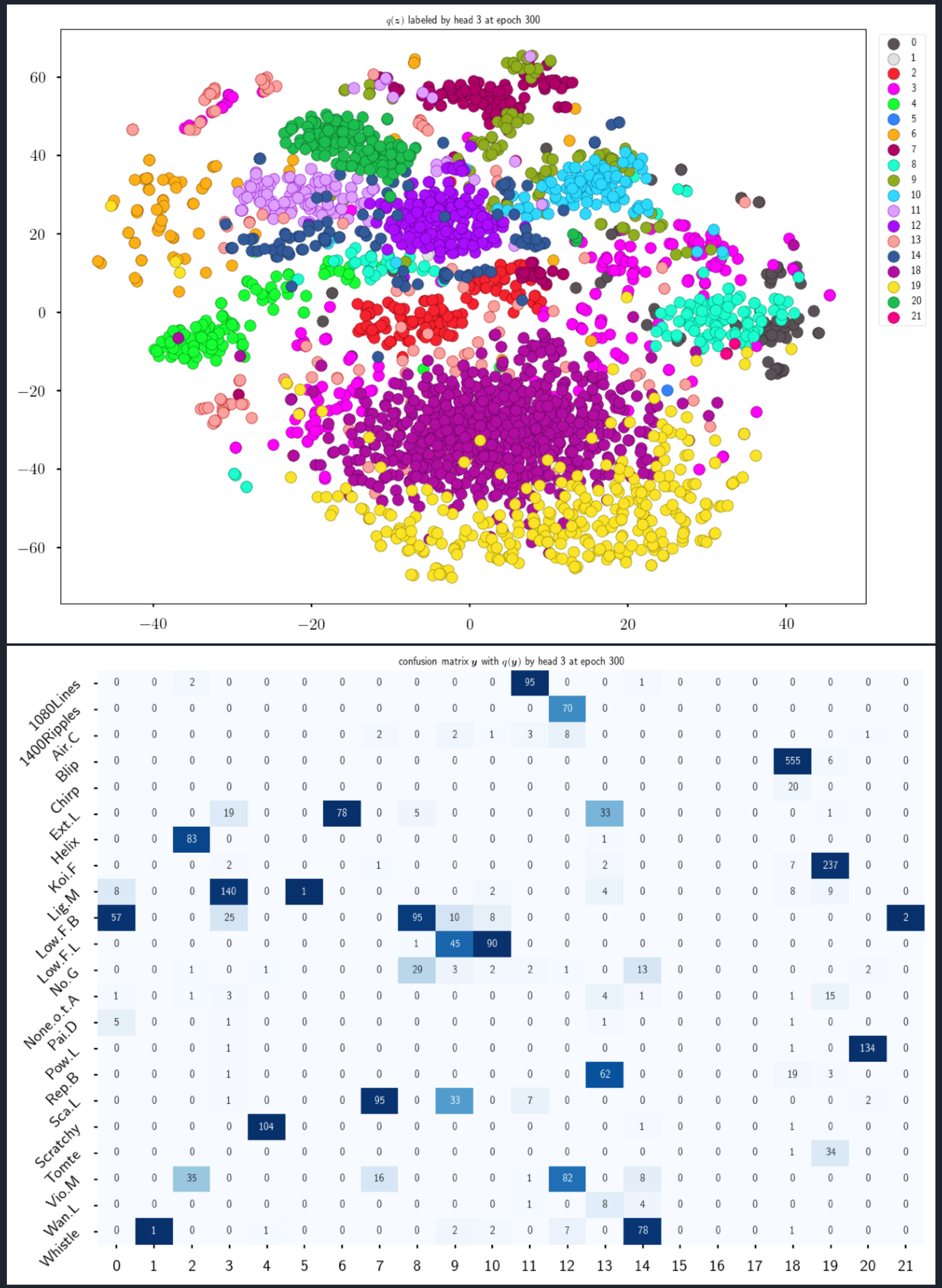
Architecture of IIC

[3] Xu Ji, et.al, "Invariant Information Clustering for Unsupervised Image Classification and Segmentation", arXiv:1807.06653v4 (2019).

Result-2: Classify features by IIC



(a) classifier-0



(b) classifier-1

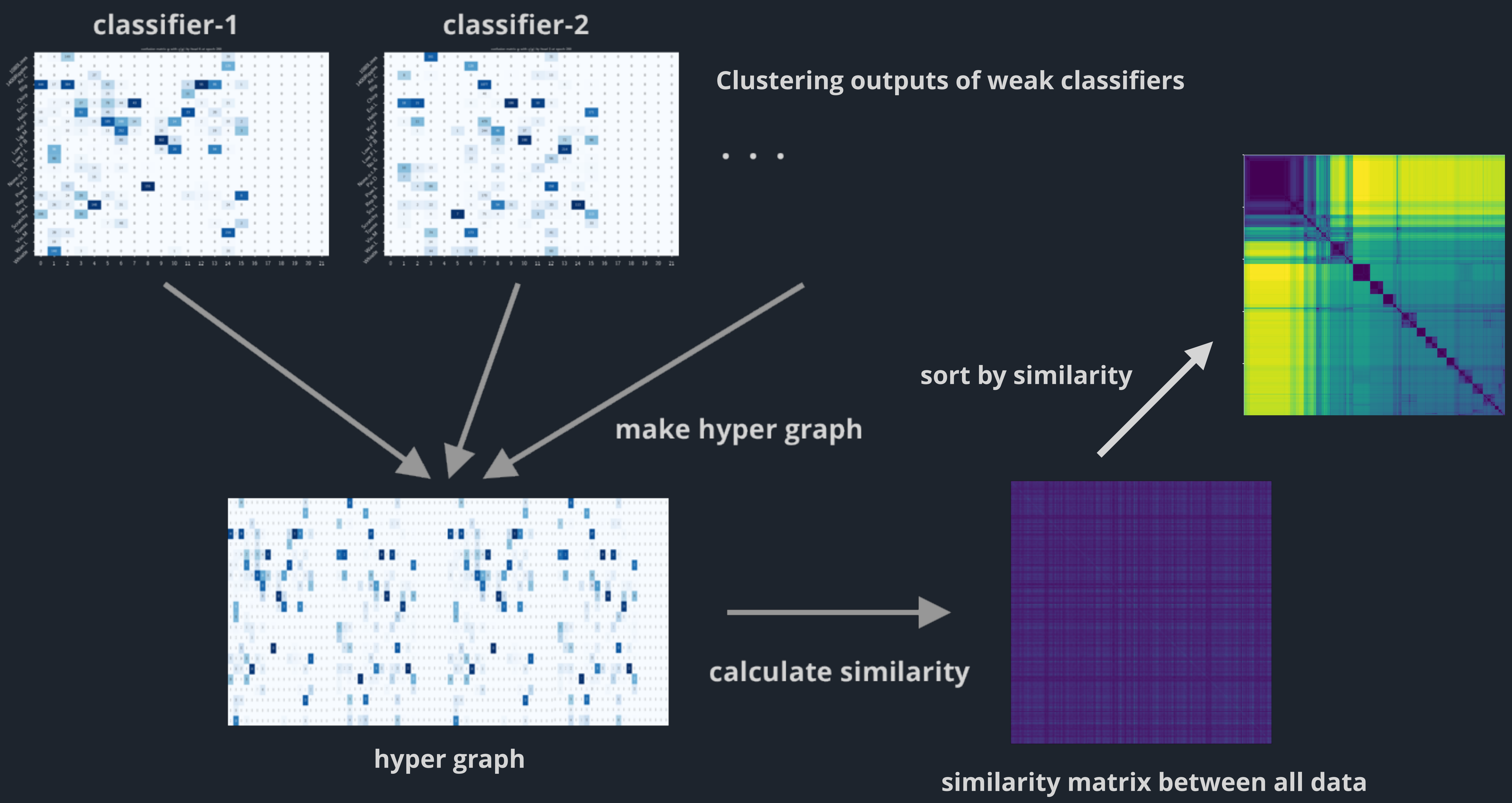
2D latent features with classifiers (upper), confusion matrix (lower)

Consideration

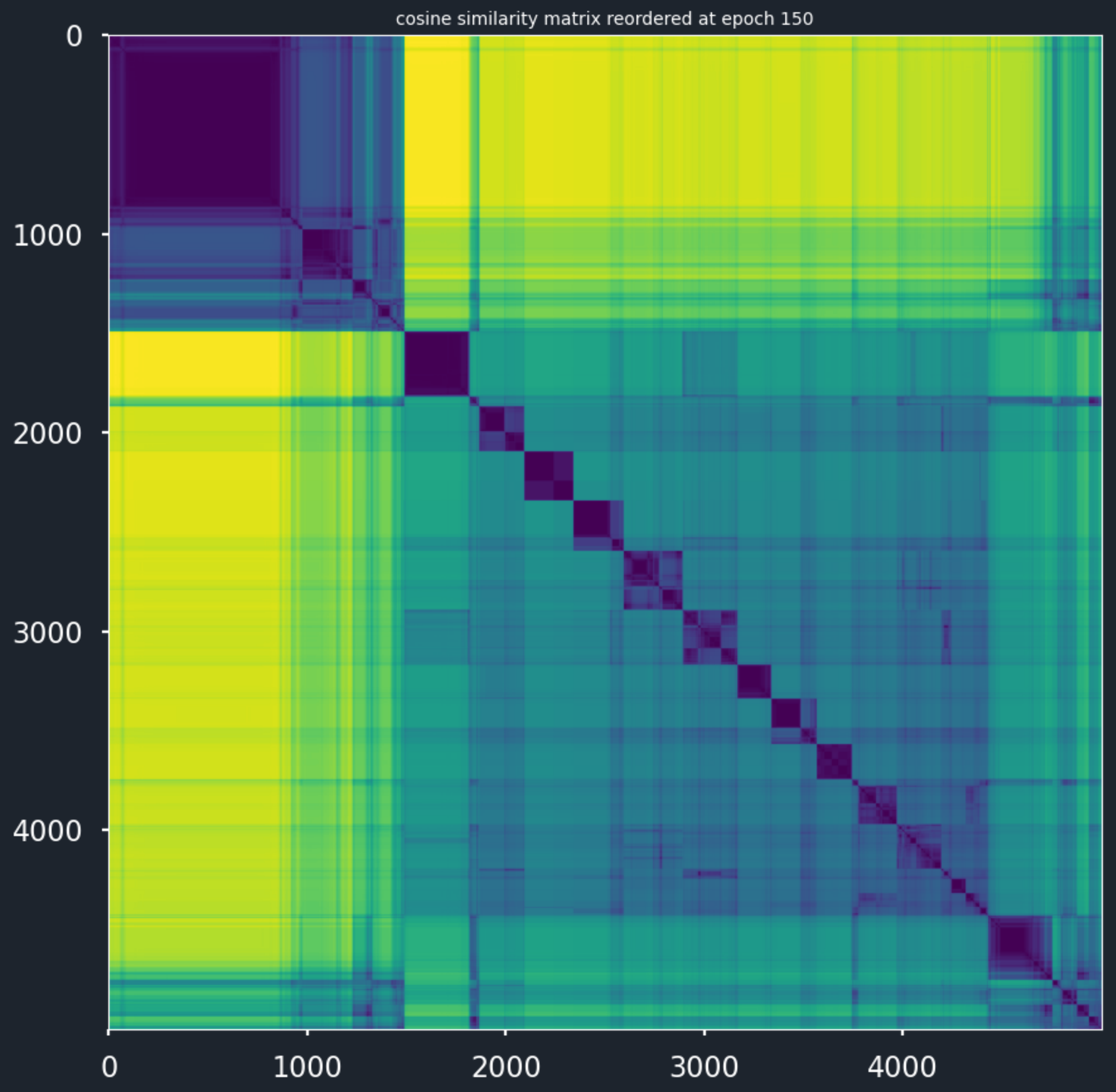
- Each classifier has its own domain for good classification.
- Clustering works well even to a certain extent with imbalanced data.
- Classification performance of classes with extremely small numbers of data is poor (e.g. Chirp).

Ensemble (preliminary)

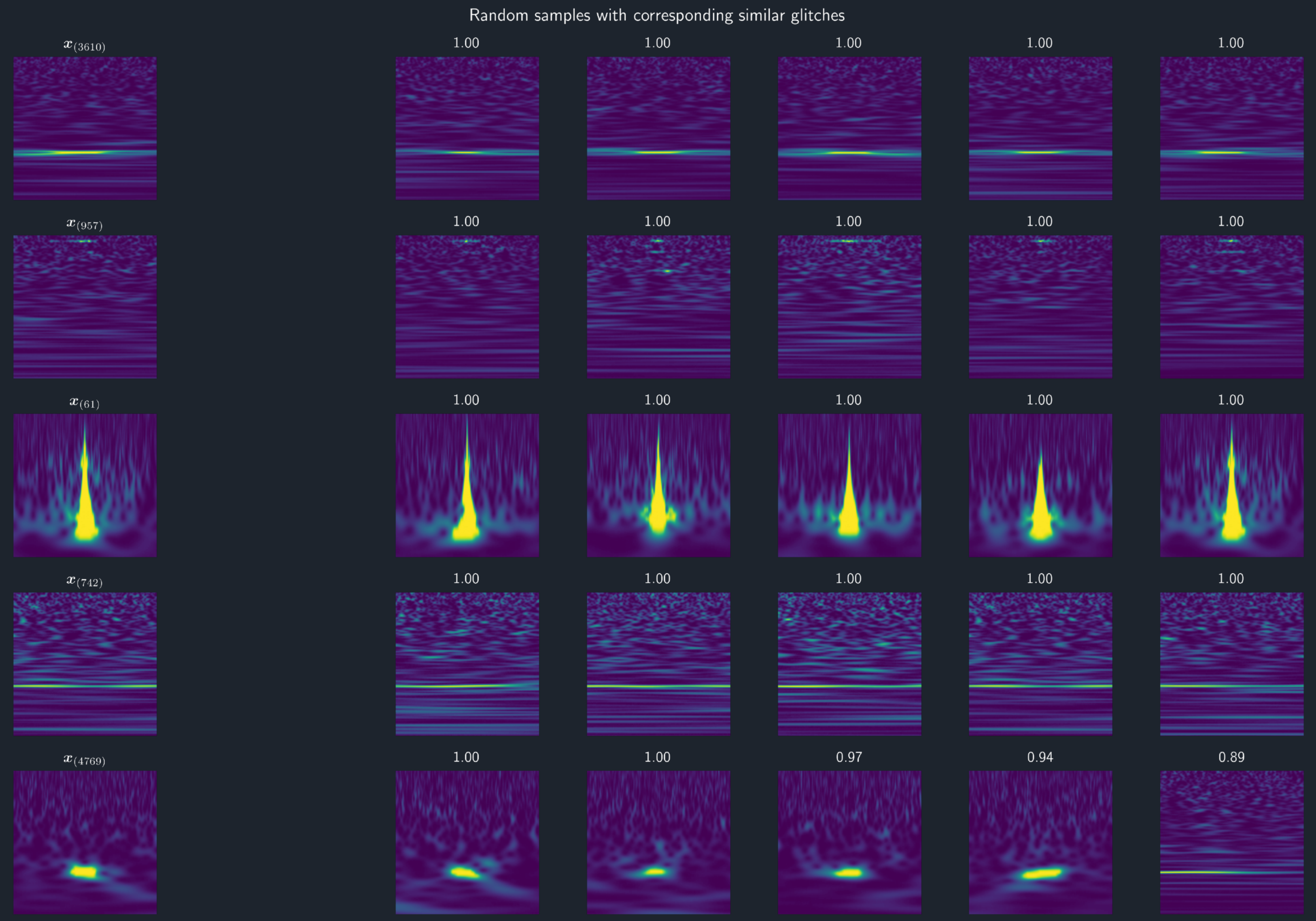
- Concatenate outputs from all weak classifiers to make **hyper graph**
- Calculate similarity matrix between all test data



Result-3: Ensemble outputs of all weak classifiers (preliminary)



Sorted cosine distance matrix of hyper graph (test data)



Random samples (left) with the 5 most similar samples (right)

Consideration & Future works

Consideration of results

- We could get class-dependent features from Gravity Spy Dataset by using VAE.
- We could see the different kinds of results from IIC weak classifiers.
- We could get similar glitches of any sample by ensemble all outputs from IIC weak classifiers.

Future works

- Consider methods to improve ensemble clustering.
- Apply our system to KAGRA data and evaluate its performance.