Unsupervised star-galaxy segmentation and classification

A summary report

Hao Sun[†◇], Edward J. Kim[†], Sofie Li[†*], Robert J. Brunner(Advisor)[†]

[†]LCDM, University of Illinois at Urbana Champaign

[◇]Peking University

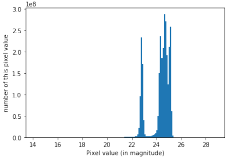
[*]Fudan University

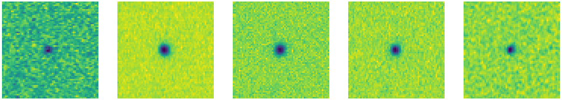
September 14, 2017

Contents

Labeled SDSS dataset

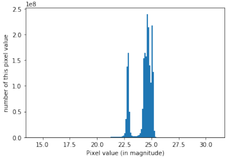
- The first dataset is labeled SDSS dataset
  1. 20,000 *.npy files, each may contain different numbers of object images (1 or more). Total 140,000 images available.
  2. x[0].dtype.names = ('objID', 'image', 'class', 'z')
  3. x[0]['image'].shape = (5, 64, 64), 5 channels are 0th -> u, 1st -> g, 2nd -> r, 3rd -> i, 4th -> z
  4. x[0]['class'] = u'STAR'
- Histograph of pixel values and some images before normalization

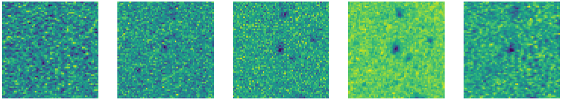




Unlabeled SDSS dataset

- The second dataset is unlabeled SDSS dataset
  1. 100,000 *.npy files, each contains 1 object image.
  2. x.dtype.names = ('image')
  3. x['image'].shape = (5, 64, 64), 5 channels are
     0th -> u, 1st -> g, 2nd -> r, 3rd -> i, 4th -> z
- Histograph of pixel values and some images before normalization





Pixel-wise segmentation: AE + Hypercolumns

- Feature learning for segmentation
  In the segmentation task, we don't aim at using any point in the hidden space to generate new images. To obtain better performance, residual connections are used.

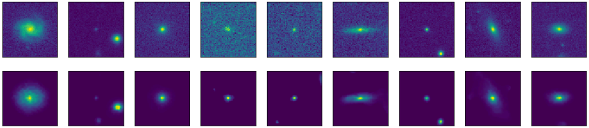


Figure: Reproduction(secondline) of input images(first line) using AE

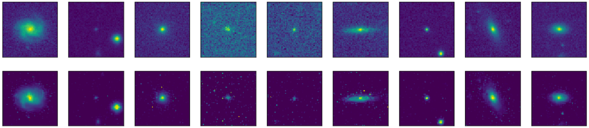


Figure: Reproduction(secondline) of input images(first line) using residual connections

Pixel-wise segmentation: AE + Hypercolumns

- Hypercolumns
  We use hypercolumns to combine the features.[1]

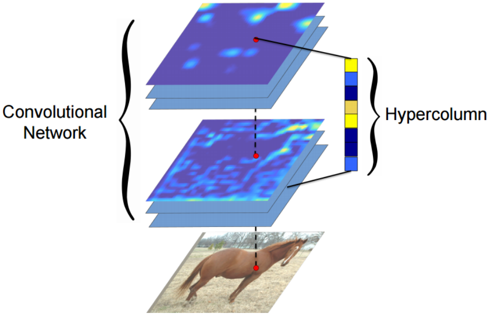


Figure: Hypercolumn representation (Image source: [Hariharan et al. (2014)])

[1] Hypercolumns for Object Segmentation and Fine-grained Localization by Hariharan et al.

Pixel-wise segmentation: AE + Hypercolumns

- Hypercolumns + clustering
  Using k-means, we can get result with different resolution power. (Trade-off)

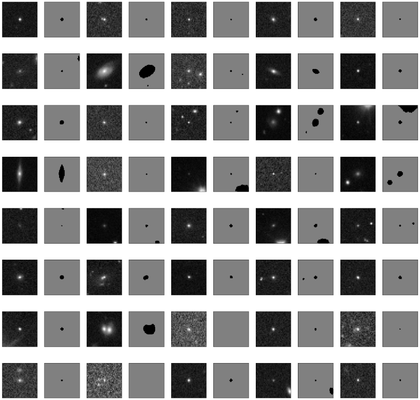


Figure: Segmentation result without residual connection

Segmentation for Classification

There may exist multiple objects in each image, which is not conducive to classification task. Sofie is now focusing on using the segmentation result to improve classification performance.

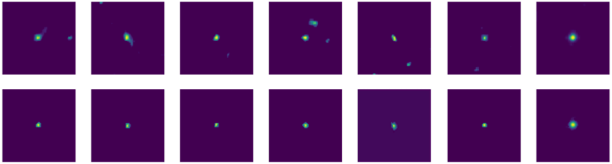


Figure: Segmentation result can help the classification part

Unsupervised classification

In the classification task, we hope we can make use of VAEs'dimension reduction ability to make clustering easier. And we try to use different clustering algorithms to separate the hidden variables of VAEs in the hidden space. However, VAEs are not designed to be a classifier but a generator. In order to get better classification performance, we tried different approaches:

1. use traditional VAEs and Manifold learning algorithms and set the result as a baseline
2. revise the loss function of VAEs
3. use a hierarchical structure which uses an AE to extract low-level features and uses a VAE to do classification.

Baseline: VAE + Manifold learning

The first idea is to use the commonly used method of unsupervised clustering. This method has three steps.

1. use VAEs or AEs to extract features from original images and reduce the dimensions into d_{hid}, where $d_{hid} \ll D_{input}$
2. run a manifold learning algorithm(such as ISOMAP, t-SNE etc.) to map the d dimensional d_{hid} into a 1 dimensional scalar
3. use the scalar and a threshold(can be determined by prior knowledge like star-galaxy proportion) to perform classification

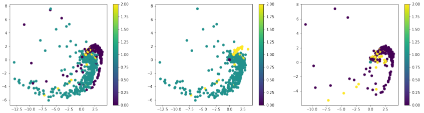


Figure: Hidden variables of VAE stars are purple and galaxies are green

AUC $= 0.75$

Improvement: Revise the KL-term

- some preliminaries for VAEs

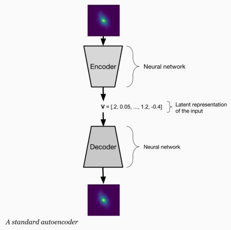


Figure: structure of AE

In AEs, there are always two parts: the encoder and the decoder(as is shown in Fig.1), which can be defined as transitions ϕ and ψ, such that

$$\phi : \mathcal{X} \rightarrow \mathcal{H} \tag{1}$$

$$\psi : \mathcal{H} \rightarrow \mathcal{X} \tag{2}$$

$$\phi, \psi = \arg\min_{\phi,\psi} \parallel X - (\phi \circ \psi) X \parallel^2 \tag{3}$$

Improvement: Revise the KL-term

- some preliminaries for VAEs

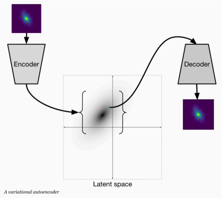


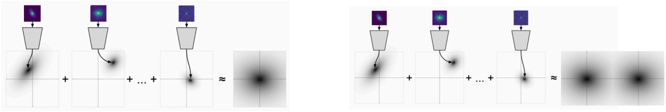
Figure: structure of AE

For the Encoder part of the VAEs, a certain class of input images are mapped to a certain Multi-dimensional, depending on the number of hidden variables, Gaussian distribution, as is shown in Figure 1(b). And then the Decoder uses a resampled hidden value to generate a new image.

Improvement: Revise the KL-term

- some preliminaries for VAEs

$$L_{VAE} = \| X_{in} - X_{out} \|^2 + D_{KL}(P(V), \mathcal{N}(0, I)) \tag{4}$$

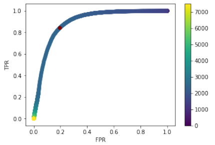


(a) prior $= \mathcal{N}(0, I)$ (b) prior $= \frac{1}{2}N(-m, s^2) + \frac{1}{2}N(m, s^2)$

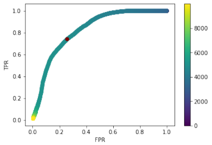
Figure: Hidden distribution with different priors

Sometimes when we train the neural net with a KL-divergence or JS-divergence, we meet the gradient disappearence problem. And this is exactly the problem people met when training Generative Adversarial Networks(GANs). We also proposed two analogies for Wasserstein metric, called AW and PW.

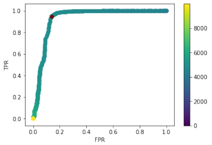
Improvement: Revise the KL-term



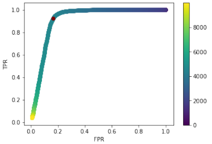
(a) V+M: AUC=0.89



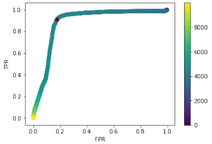
(b) Vr(SC)+M: AUC=0.82



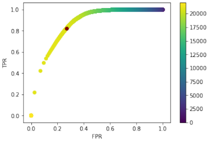
(c) Vr(AW)+M: AUC=0.92



(d) Vr(DKL)+M:AUC=0.91



(e) Vr(PW)+M: AUC=0.88



(f) A+Vr(KL): AUC=0.89

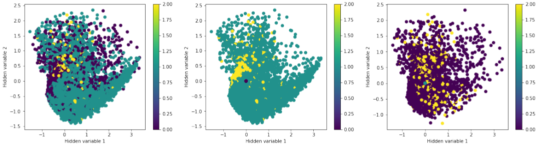
Figure: ROC and AUC of each method

Problems: Instability. AUC in repeat experiment varies from 0.6 to 0.92

Solution: A robust AE + Conditional VAE design

The most successful generative models often use only a single layer of latent variables [2], and those that use multiple layers only show modest performance increases in quantitative metrics such as log-likelihood [3].

1. when using more hidden variables (30), the representation ability would be more sufficient to generate consistant images. But the features are more likely to be in lower-level. So that unsupervised clustering methods can not work here.
2. when using less hidden variables (2), the representation ability would be less sufficient to generate consistant images, although the features are more likely to be in higher-level. So that accuracy of unsupervised learning is limited.



[2]Radford et al., 2015; van den Oord et al., 2016
[3]Sønderby et al., 2016; Bachman, 2016

A hierarchical design

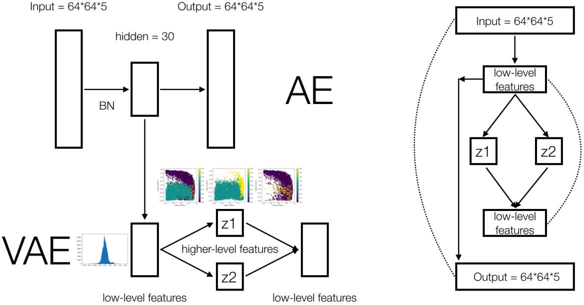


Figure: Left:AE + VAE; Right: VAE with the loss defined by AE

Conditional VAE

In Conditional Variational Auto-Encoders(CVAEs), a label denotes the class which the image belongs to is also part of the input for both encoder and decoder. We come up with an analogy of CVAEs in an unsupervised way: one hidden unit is constrained by the KL-term and is used as a classifier while other units without KL-term contribute to better reproduction.

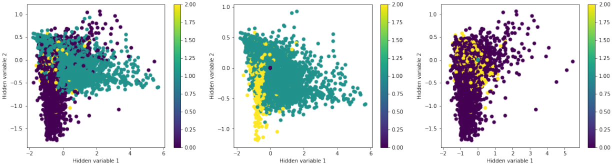


Figure: X-axis: hidden unit for classification; Y-axis: hidden unit for reproduction

Result: AUC $= 0.903$ on average in 100 experiments
with standard deviation$=0.002$