EFFICIENT UNSUP. LEARNING FOR PLANKTON IMAGES

Alfano P., Rando M., Letizia M., Rosasco L., Odone F., Pastore V. - Genoa University
{paolodidier.alfano, marco.rando, marco.letizia}@edu.unige.it, {lorenzo.rosasco, francesca.odone, vito.paolo.pastore}@unige.it

ICVSS 2022
Sicily ~ 10-16 July
International Computer Vision Summer School

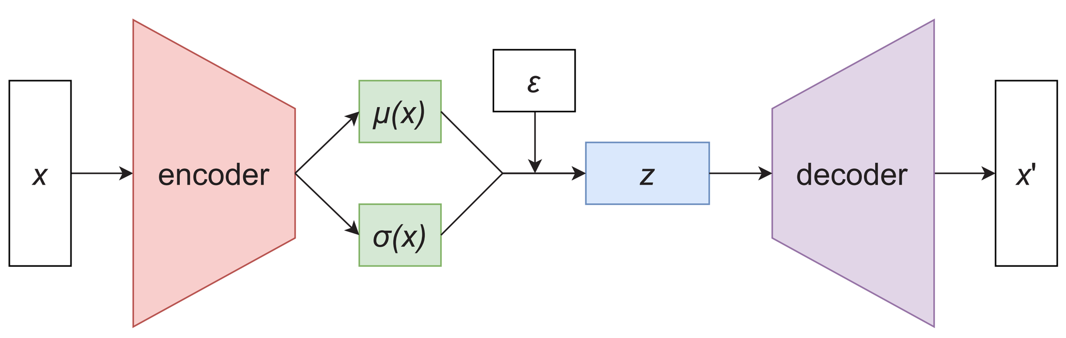
Abstract

Monitoring plankton populations, reacting to minimal changes in the environment, is fundamental to preserve the aquatic ecosystem. In this context, the adoption of machine learning algorithms may be affected by the significant cost of manual annotation. To address these challenges, we propose an efficient unsupervised learning pipeline. First, a Variational Autoencoder is trained on features extracted by a pre-trained neural network. Then we use the learnt latent space for clustering.

Methodology

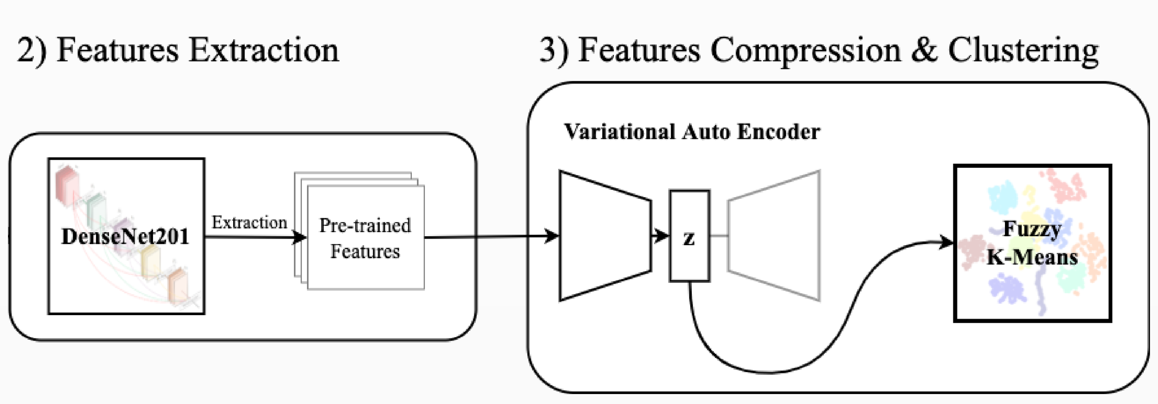
A Variational Autoencoder(VAE)[2] replicates input while learning a lower dimensional embedding by encoding the input into a latent distribution:

$$x' = d(e(x))$$
$$l(x, x') = \left\|x - x'\right\|^2 + D_{KL}(\mathcal{G}, \mathcal{N}).$$



Our pipeline:

1. Image pre-processing
2. Features extraction: a neural network pre-trained on ImageNet produces output features.
3. Features compression & clustering: train a convolutional Variational Autoencoder. Learnt latent space is fed to a clustering algorithm.



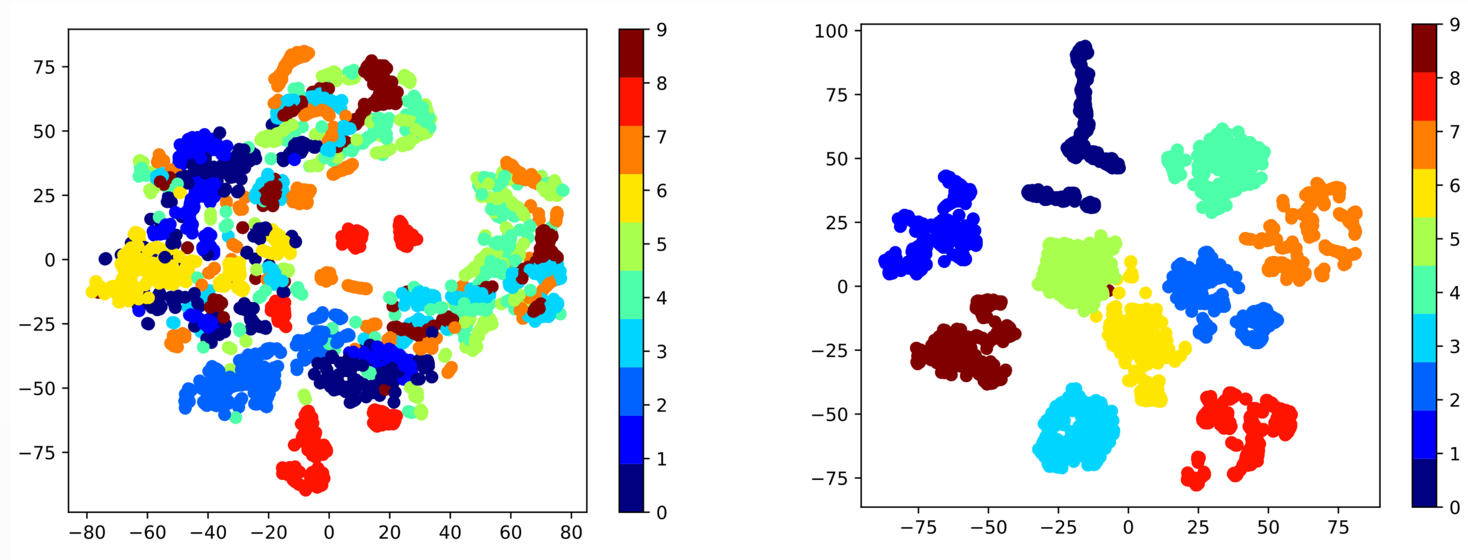
Datasets & Experiments

- Lensless: 10 classes, 640 color images each.
- WHOI40: 40 classes, 100 grayscale images each.
- WHOI22: 22 fine grained species, 300 grayscale images each.

We evaluate our pipeline with the *purity* measure:

$$\mathrm{purity}(\Omega, C) = \frac{1}{N}\sum_k \max_j |w_k \cap c_j|$$

$\Omega = \{w_1, ..., w_k\}$ set of clusters, $C = \{c_1, ..., c_j\}$ set of real classes. We show qualitative and quantitative benefit using pre-trained features on the lensless dataset.



Algorithm/Z	10	30	50	100	500
image-VAE	0.53 ± 0.017	0.55 ± 0.04	0.58 ± 0.01	0.59 ± 0.01	0.62 ± 0.01
FE-VAE	0.98 ± 0.01	0.98 ± 0.03	0.98 ± 0.01	0.98 ± 0.02	0.98 ± 0.02

- Pre-trained features increases the purity by 30% on average
- Using a bigger latent space does not imply a performance improvement

Quantitative results on the other datasets:

Dataset/Z	10	30	50	100	500
WHOI 40	0.66 ± 0.01	0.71 ± 0.02	0.73 ± 0.02	0.77 ± 0.01	0.77 ± 0.01
WHOI 22	0.63 ± 0.004	0.66 ± 0.01	0.68 ± 0.005	0.68 ± 0.006	0.68 ± 0.01

- Best performances correspond to a latent space size $Z = 100$.

We compared our results with a state-of-the-art unsupervised learning pipeline.

Algorithm/Dataset	Lensless	WHOI 40	WHOI 22
Pipeline from [3]	0.93	0.71	0.56
Ours	**0.98**	**0.77**	**0.68**

Conclusions

We introduced an efficient unsupervised pipeline for plankton images. Input images are fed to a pre-trained neural network. Output features are used as inputs to train a Variational Autoencoder. The latent space representation is used by clustering algorithm. Future developments will extend our analysis to other datasets to test our pipeline on a more general context.

References

[1] Boyce, D et al. Global phytoplankton decline over the past century, in *Nature 466*, 2010
[2] Diederik, K et al. Auto-Encoding Variational Bayes, in *arXiv*:1312.6114, 2014
[3] Pastore, V et al. Annotation-free learning of plankton for classification and anomaly detection, in *Scientific Reports 10*, 2020