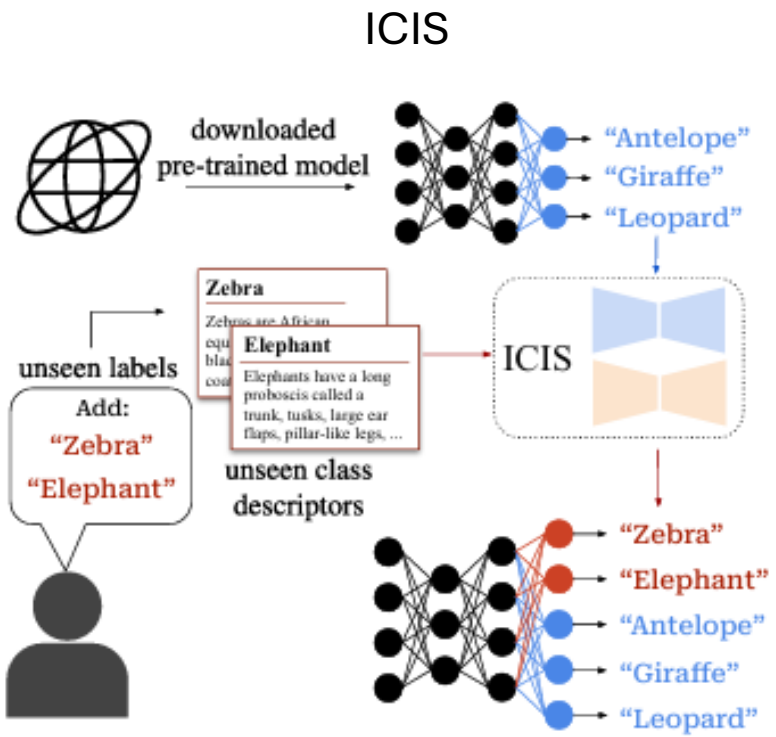
UNIVERSITÀ DI PISA

Thesis presentation

A.Y. 2023/24

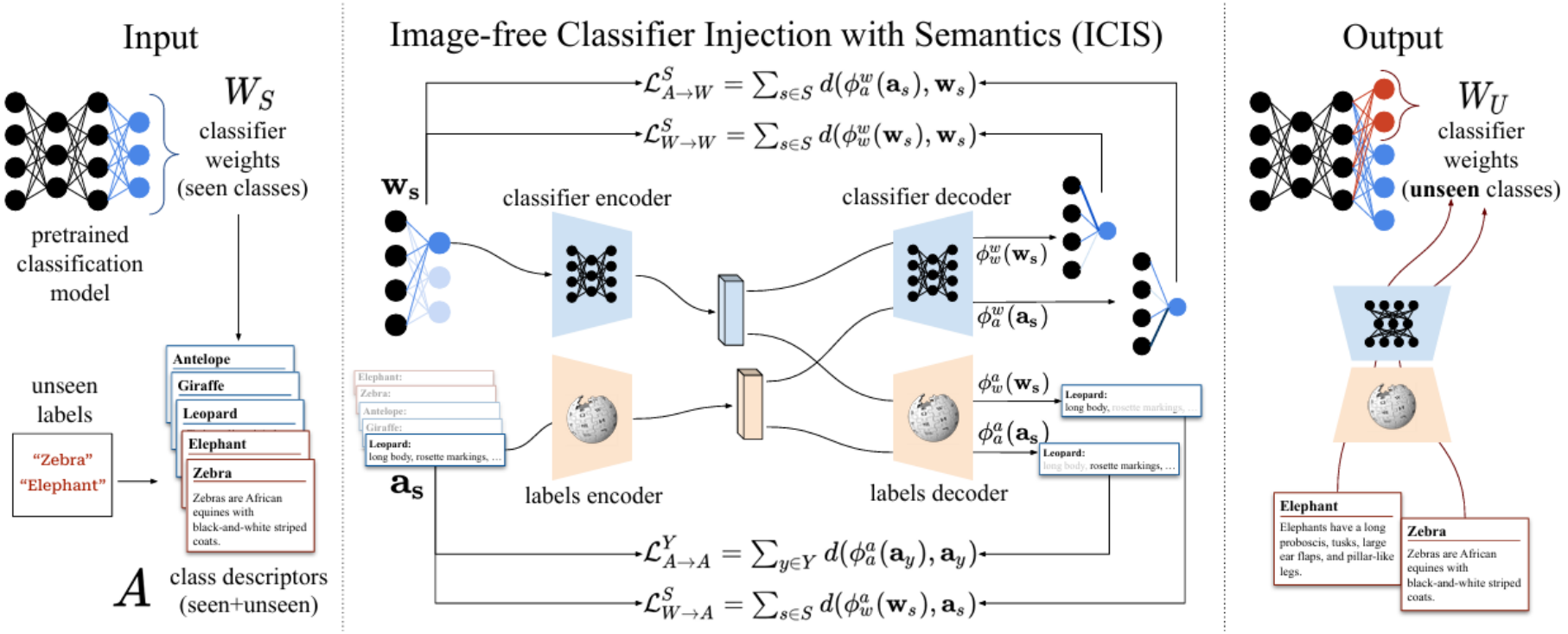
Federico Frati

Visual Embedding Representations for Zero-Shot Learning in Computer Vision



Introduction:
Image free zero shot learning

- Given a specific image classification task and a pre-trained model, can we extend it to desired but missing categories without using images from seen or unseen classes?
- Do the classifiers' weights implicitely bring information about images?
- Is it possible to generalize the problem to other task types? E.g. clustering



A proposed architecture for zero-shot images classification: ICIS

- Loss as the sum of four different losses
- Encoder decoder based
- Completely image free
- Good performances compared to pre-existing methods in the literature applicable to image-free ZSL

Image-free Zero-Shot Learning	Zero-Shot Accuracy			Generalised Zero-Shot Accuracy								
	CUB	AWA2	SUN	CUB			AWA2			SUN		
	Acc	Acc	Acc	u	s	H	u	s	H	u	s	H
ConSE [39]	41.9	44.0	44.4	0.5	88.0	0.9	3.0	96.1	5.7	0.1	47.9	0.1
COSTA [36]	31.9	40.9	19.9	0.0	87.6	0.0	0.0	96.1	0.0	0.0	50.1	0.0
Sub. Reg.* [2]	37.6	37.5	48.3	0.0	87.6	0.0	0.0	96.1	0.0	0.0	50.1	0.0
wDAE* [16]	38.2	37.0	49.9	0.0	87.3	0.0	0.1	96.0	0.3	0.0	49.3	0.0
WAvg* [61]	2.0	20.1	1.4	1.9	52.3	3.7	5.5	92.4	10.4	0.0	50.1	0.0
SMO* [61]	45.1	55.4	42.7	39.2	52.3	44.8	31.8	92.4	47.3	42.5	1.6	3.1
ICIS (Ours)	**60.6**	**64.6**	**51.8**	45.8	73.7	**56.5**	35.6	93.3	**51.6**	45.2	25.6	**32.7**

Experimental results

- CUB: fine-grained bird classification
- AWA2: coarse-grained classification with 50 different animals
- SUN: dataset for indoor and outdoor scenes classification

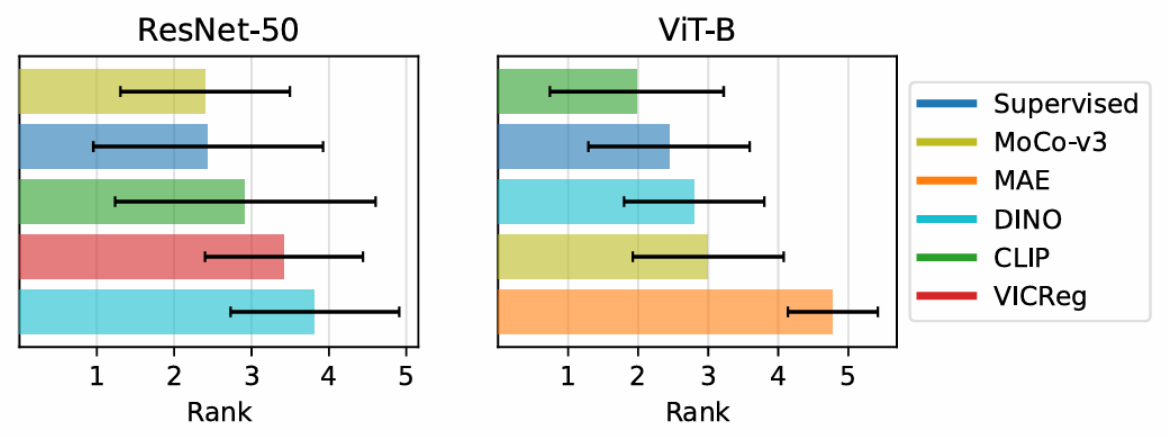
What about clustering?

$$\mathrm{AMI}(U,V) = \frac{\mathrm{MI}(U,V) - \mathbb{E}[\mathrm{MI}(U,V)]}{\mathrm{mean}(\mathrm{H}(U) + \mathrm{H}(V)) - \mathbb{E}[\mathrm{MI}(U,V)]}$$

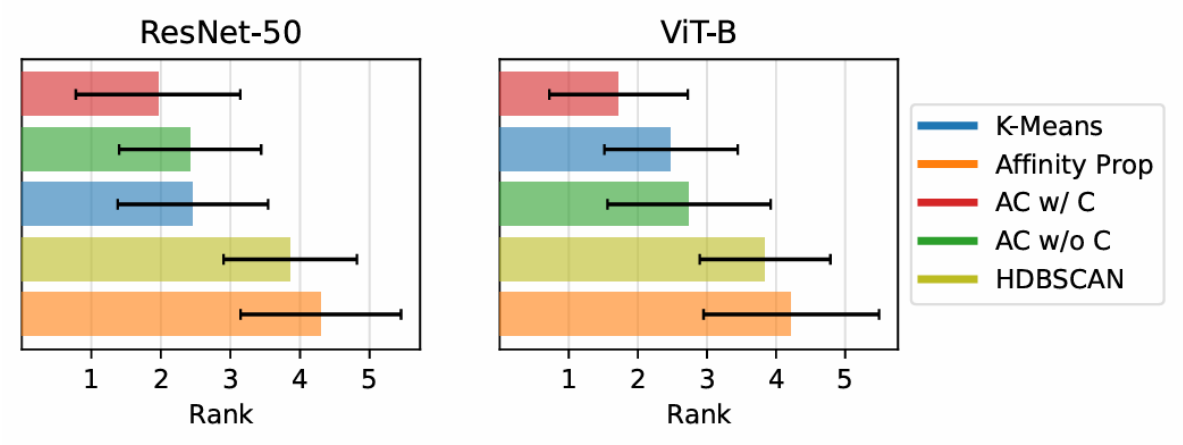
$$S = \frac{1}{N}\sum_{i}^{N}\frac{a_i - b_i}{\max(a_i, b_i)}$$

- Self-supervised learning (SSL) has seen a large amount of interest in recent years across almost every machine learning sub-field, due to the promise of being able to harness the large quantities of unlabeled data available
- Zero-shot clustering of feature embeddings: tests considering different methods from the major self-supervised paradigms and different clustering algorithms
- Silhouette score is strongly correleted with AMI and can be considered a good indicator for performances
- AMI unusabe if there is no ground truth available

Results



Average rank of each tested SSL encoder (lower is better). For both the ResNet-50 and ViT-B backbones an SSL encoder in general results in the best clustering. It is worth noting that the supervised method also in general produces good clusters



Average clustering method rank (lower is better). AC method performs very well, whether the number of cluster are known a priori or not .

My thesis work

With this thesis work we want to try to place ourselves between the two approaches presented. The details have not yet been defined, but the key idea is to investigate and exploit the information implicitly present in the feature embeddings to perform zero shot tasks without the direct use of images in a Self Supervised approach.

UNIVERSITÀ DI PISA

Thank you for the attention

Thesis presentation

Federico Frati

A.Y. 2023/24

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE