# COMPUTER VISION
## MEI/1

**University of Beira Interior**

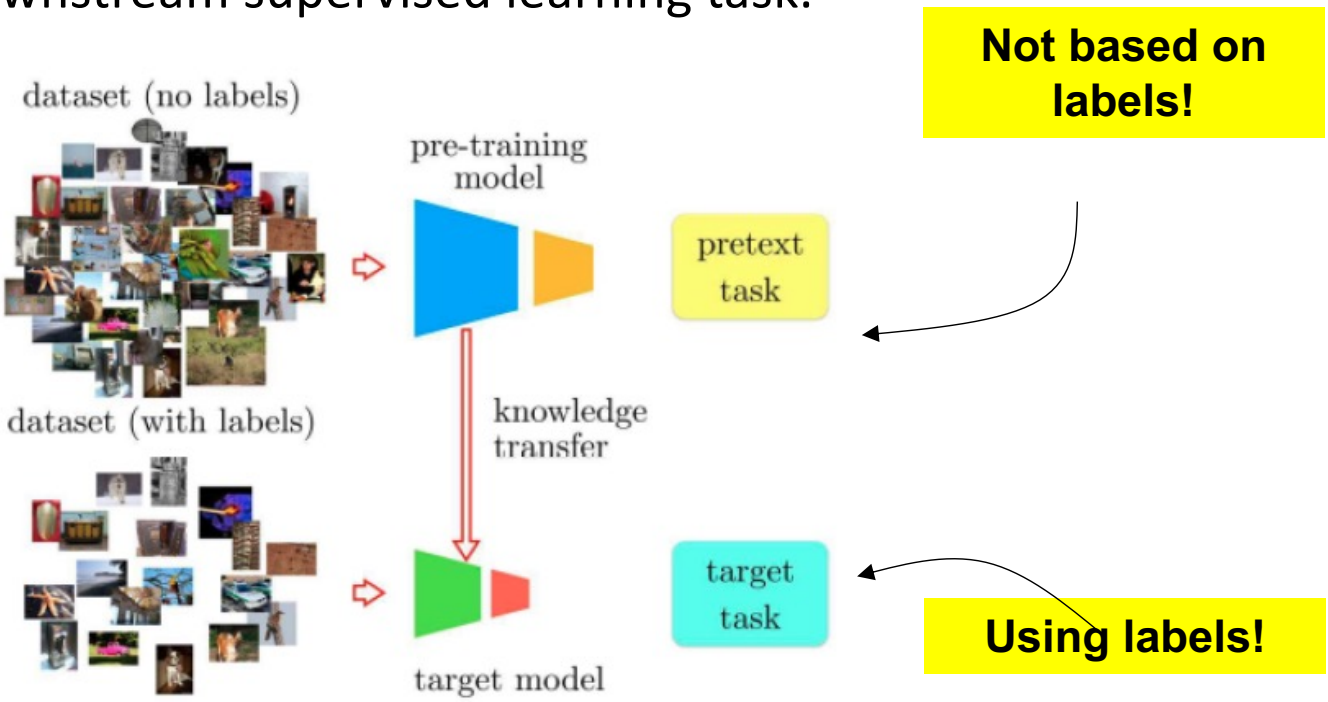
Department of Informatics

Hugo Pedro Proença

hugomcp@di.ubi.pt, **2024/25**

# Self-Supervised Learning

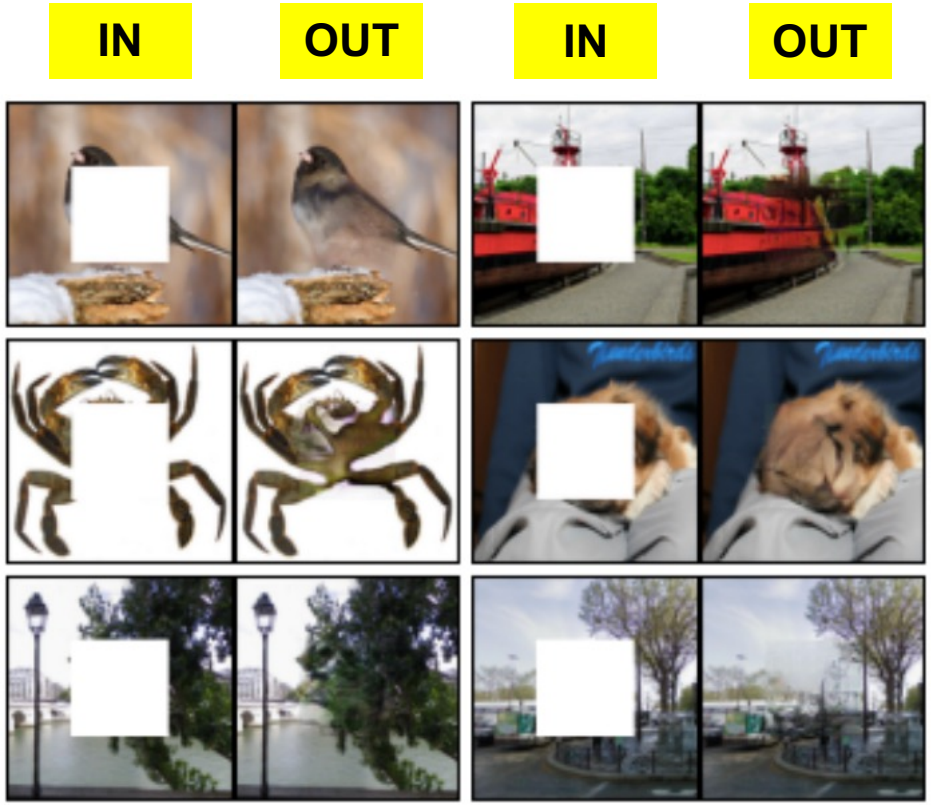
- Self-supervised learning is a recent type of machine learning that can be regarded as a midle point between supervised and unsupervised learning.
- It is a form of unsupervised learning where the model is trained on unlabeled data, but the goal is to **learn good representations** of the data that can belater used in a downstream supervised learning task.




Source: https://neptune.ai/blog/self-supervised-learning

# Self-Supervised Learning

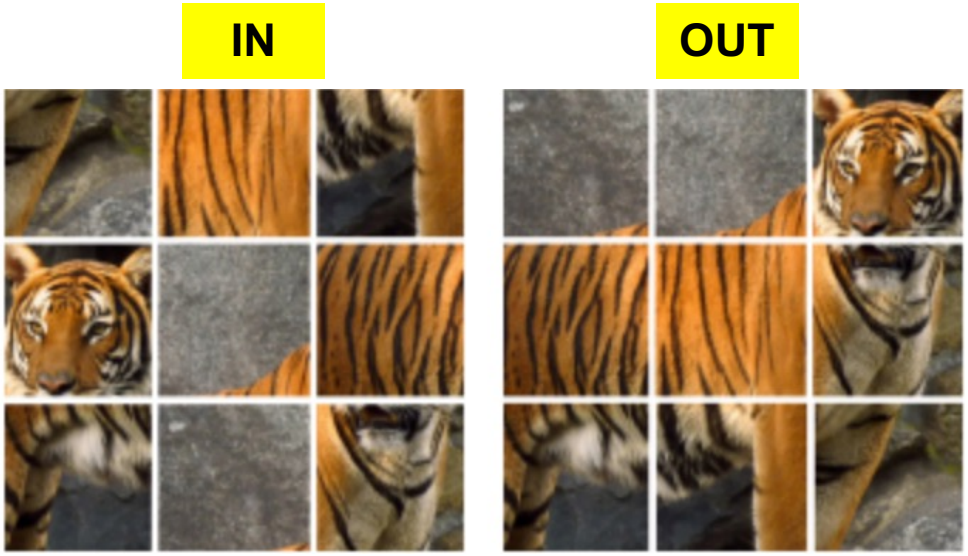
- At first, Self-supervised learning starts by training a model itself to learn one part of the input from another part of the input.
- This is known as **pretext learning**, which can assume different forms:
- For example, using unstructured 2D data, predict any part of the input from any other part:




**By doing this, we force the model to "understand" the data**

# Self-Supervised Learning

- Still for unstructured 2D data, another very popular pretext task is to learn by solving Jigsaw puzzles:

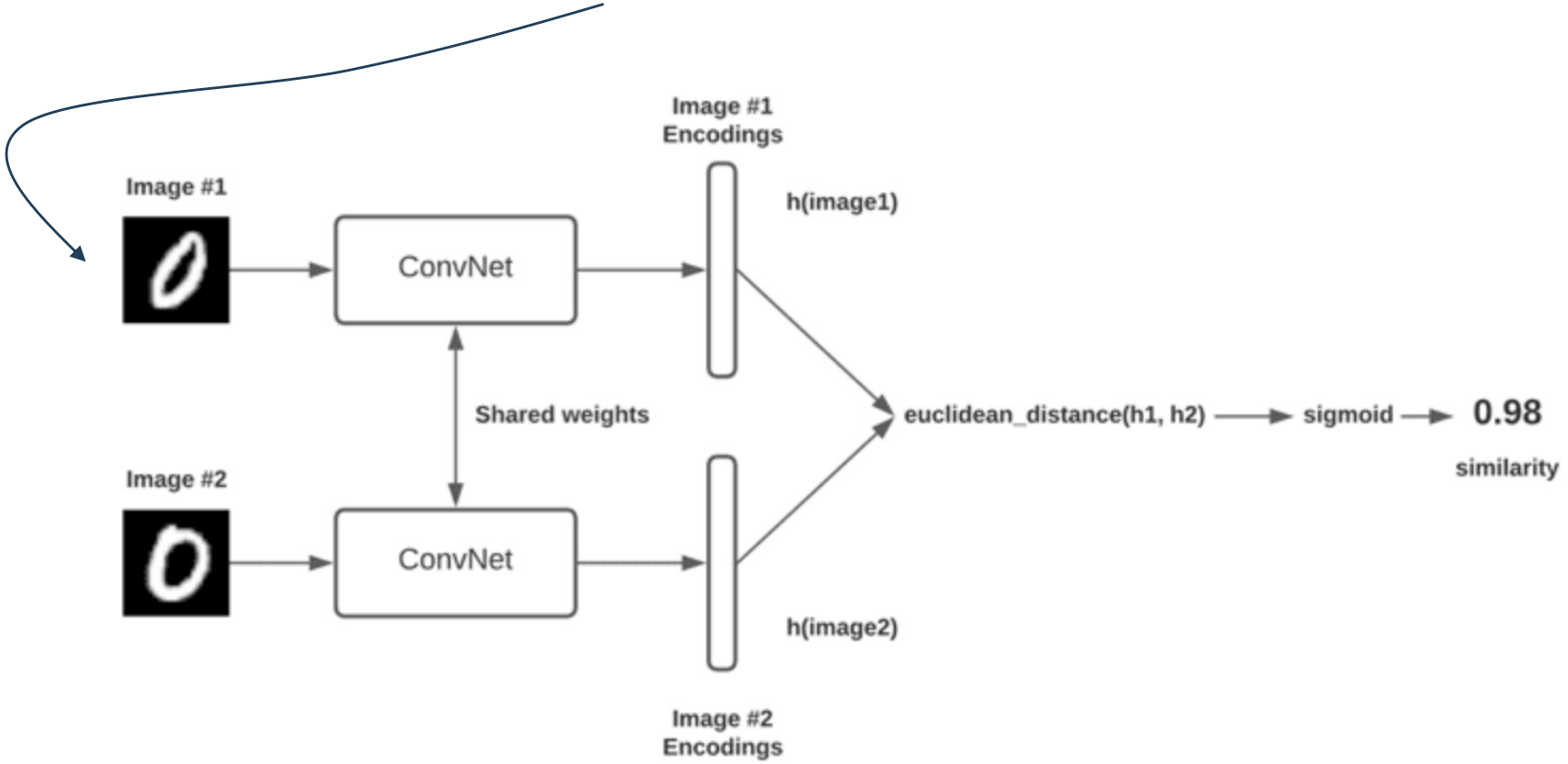



**Again, the model is forced to understand each part of the input, in order to obtain a realistic output**

# Self-Supervised Learning

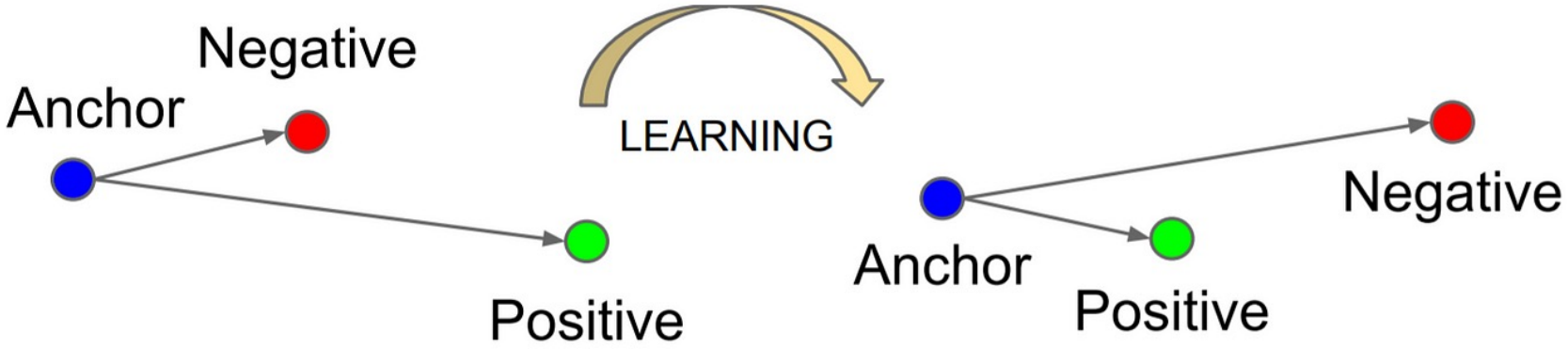
- It is also very common to use some Siamese architecture to obtain appropriate feature representations.

**If both inputs are from the same image (not "class" in this case), the distance should be small. Otherwise, it should be large.**



# Self-Supervised Learning

- Another possibility is to use three images in the input: the Anchor (**A**) and the Positive (**P**) that are variations of the same image, and the negative (**N**), that regards a different image.



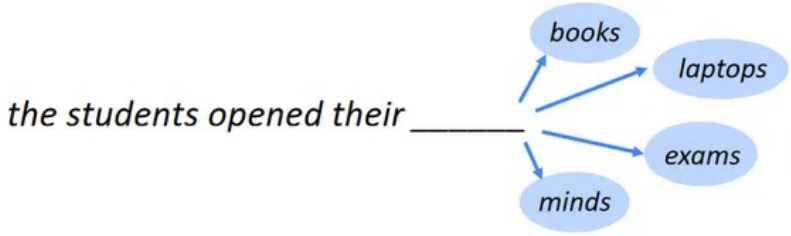

**The Anchor and Positive should be near each other, while their distance to the Negative image should be large**

$$\mathcal{L}\left(A, P, N\right) = \max(\| \mathrm{f}(A) - \mathrm{f}(P) \|_2 - \| \mathrm{f}(A) - \mathrm{f}(N) \|_2 + \alpha, 0)$$

# Self-Supervised Learning

- In case of 3D unstructured data (video), one can predict the predict the future from the past/present, or predict the present from the future.

- In case of text data, the most obvious pretext task is to predict the next word, based in the last “k” words.



# Self-Supervised Learning

- Once the pretext task is considered solved (i.e., the model stopped to learn), it is time to apply “Transfer Learning” techniques
- In practice, it consists in copying (and freezing ?) the weights from the earliest layers of the model into the new one.