

# Semantic-Based Few-Shot Learning by Interactive Psychometric Testing

Lu Yin\*, Vlado Menkovski, Yulong Pei, Mykola Pechenizkiy

Eindhoven University of Technology, Eindhoven 5600 MB, Netherlands  
 {l.yin,v.menkovski,y.pei,l.m.pechenizkiy}@tue.nl

## Abstract

Few-shot classification tasks aim to classify images in query sets based on only a few labeled examples in support sets. Most studies usually assume that each image in a task has a single and unique class association. Under these assumptions, these algorithms may not be able to identify the proper class assignment when there is no exact matching between support and query classes. For example, given a few images of lions, bikes, and apples to classify a tiger. However, in a more general setting, we could consider the higher-level concept, the large carnivores, to match the tiger to the lion for semantic classification. Existing studies rarely considered this situation due to the incompatibility of label-based supervision with complex conception relationships. In this work, we advance the few-shot learning towards this more challenging scenario, the semantic-based few-shot learning, and propose a method to address the paradigm by capturing the inner semantic relationships using interactive psychometric learning. The experiment results on the CIFAR-100 dataset show the superiority of our method for the semantic-based few-shot learning compared to the baseline.

## Introduction

With enormous amounts of labeled data, deep learning methods have achieved impressive breakthroughs in various tasks. However, the need for large quantities of labeled samples is still a bottleneck in many real-world problems. For this reason, few-shot learning (Lake et al. 2011; Vinyals et al. 2016) is proposed to emulate this by learning the transferable knowledge from the “base” dataset where ample labeled samples are available to generalize to another “novel” dataset which has very few labeled training examples. A popular approach for this problem is meta-learning based phase (Snell, Swersky, and Zemel 2017; Finn, Abbeel, and Levine 2017) which follows the episodic training procedure to mimic the few-shot tasks. In each few-shot task, a few labeled examples (the support set) are given to predict classes for the unlabeled samples (the query set).

While these formulations have made significant progress, the underlying assumption is that each data point from the

\*Accepted at the AAAI-22 Workshop on Interactive Machine Learning (IML@AAAI’22)  
 Copyright © 2022, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

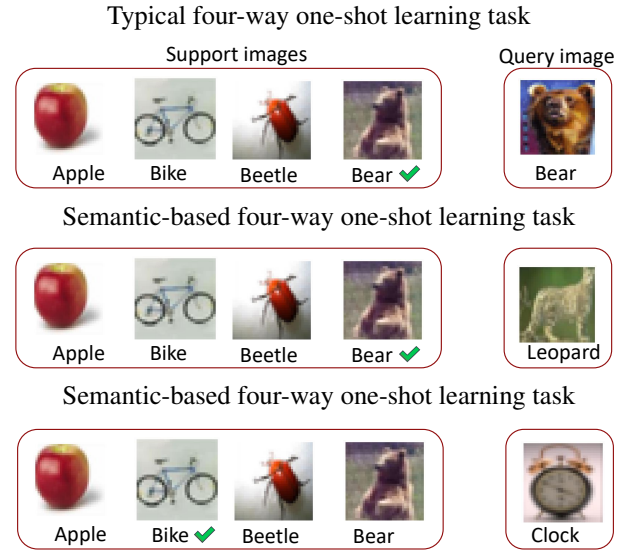


Figure 1: Different settings of few-shot learning tasks. The first row follows a typical four-way one-shot learning setting. The class of the query image matches one of the support set labels. In the second task, the typical few-shot learning model might fail to identify the query image when there is no exact class matching. However, if we could consider the higher-level semantic concept of the carnivores, a correct assignment could still be made by matching bear to leopard. A similar prediction could also be made if we consider the concept of non-living things in the last task.

support set and query set has a single and uniquely identified class association, and the query image must precisely match one of the support set classes. However, as illustrated in the last two rows in Figure 1, the few-shot learning models that are capable of dealing with classification based on the predefined classes may not be able to identify the right class assignment when there is no exact class matching.

In a more general setting, if considering the concept at a higher level, e.g., whether they are large carnivores or living things in Figure 1, one could determine the right class association. Humans are very capable of inferring these concepts on a higher level, while typical few-shot learning algorithms are not specifically designed for this under single discriminating class descriptions. They treat each class

equally without considering their intra hierarchical semantic relationships. One possible reason for this limitation might be the supervision approach: the traditional label-based supervision is incompatible with the complex conception hierarchy. Fortunately, much progress has been made in learning from other types of supervision, such as psychometric testing (Gescheider 2013). While label-based supervision reduces the comprehensive semantic relationships to given discrete labels, these psychometric testing based methods could elicit the relative conception similarities and full-depth of knowledge by transmitting the annotations progress to pair or triplet comparisons. Then the elicited knowledge could be used for other downstream tasks such as clustering or segmentation (Yin et al. 2021; Yin, Menkovski, and Pechenizkiy 2020). Enabled with such techniques, our work aims to extend the capabilities of few-shot learning models towards a more challenging setting, the semantic-based few-shot learning.

To be specific, we assume there is a shared concept hierarchy covering both base and novel classes. Self-supervised learning (SSL) is applied for feature learning at the first stage. The interactive psychometric testing is then followed to capture the similarities of the semantic concepts from base dataset. We use these semantic similarities to fine-tune the learned features from SSL, and map them to a semantic embedding space where we transfer the learned hierarchical knowledge from base classes to novel classes for semantic few-shot prediction.

Our contributions could be summarized as follows.

- \* We define a new problem setting, the semantic-based few-shot learning. It aims to identify the correct assignment to query image by higher-level concepts when there is no class matching between query and support images.
- \* We analyze the limitations of label-based supervision under the semantic-based few-shot learning setting and propose a psychometric learning based approach to tackle this problem.
- \* We evaluate our method by comparing it with a typical few-shot baseline (prototype network (Snell, Swersky, and Zemel 2017)) on CIFAR-100 dataset (Krizhevsky, Hinton et al. 2009). The results demonstrate that our method could significantly outperform this baseline in semantic-based few-shot learning even using fewer annotations from base data.

## Related Work

There are three lines of research closely related to our work: psychometric testing, few-shot learning, and self-supervised learning.

**Psychometric testing** Psychometric testing (Gescheider 2013) aims to study the perceptual processes under measurable psychical stimuli such as tones with different intensity or lights with various brightness. In general, two types of psychometric experiments could be carried. Firstly, the absolute threshold based method tries to detect the point of stimulus intensity that could be noticed by a participant. For example, how many hairs are touched to the back of hand

before a participant could notice. Secondly, the discriminative based experiments aim to find the slightest difference between two stimuli that a participant could perceive. Participants might be asked to describe the difference in direction or magnitude between these two stimuli or forced to choose between the stimuli concerning a specific parameter of interest (also known as two-alternative-force choice (2AFC) test (Fechner 1860)). Some scholars extend the 2AFC to M-AFC methods (DeCarlo 2012) by comparing  $M$  stimuli in one test to elicit the subjects' perception of more complex multimedia such as videos or images (Son et al. 2006; Feng, Marcellin, and Bilgin 2014; Yin et al. 2021; Yin, Menkovski, and Pechenizkiy 2020). In our work, we take advantage of the 3-AFC method to align with our loss function. Three samples are presented in one test to elicit the annotator's perception regarding the conception similarity.

**Few-Shot Learning** Meta-learning (learning to learn) has gained increasing attention in the machine learning community, and one of its well-known applications is few-shot learning. Three main approaches have emerged to solve this problem. Metric learning based methods aim to learn a shared metric in feature space for few-shot prediction, such as prototypical network (Snell, Swersky, and Zemel 2017), relation networks (Sung et al. 2018) and matching networks (Vinyals et al. 2016). Optimization based methods follow the idea of modifying the gradient-based optimization to adapt to novel tasks (Nichol and Schulman 2018; Finn, Abbeel, and Levine 2017; Gidaris and Komodakis 2018). Memory based approaches (Finn, Abbeel, and Levine 2017; He et al. 2020a) adopt extra memory components for novel concepts learning, and new samples could be compared to historical information in the memories.

While these frameworks lead to significant progress, little attention has been paid to leveraging the knowledge hierarchy and dealing with the situation when there is no precise label matching between query images and support images, i.e., the semantic-based few-shot learning scenario.

**Self-Supervised Learning** When human supervision is expensive to obtain, self-supervised learning could be a general framework to learn features without human annotations by solving pretext tasks. Various pretexts have been studied for learning useful image representation. For example, predicting missing parts of the input image (Trinh, Luong, and Le 2019; Larsson, Maire, and Shakhnarovich 2016; Pathak et al. 2016; Zhang, Isola, and Efros 2016, 2017), the image angle under rotation transformation (Gidaris, Singh, and Komodakis 2018), the patch location, or the number of objects (Noroozi, Pirsaviash, and Favaro 2017). Recently, another line of researches follows the paradigm of contrastive learning (Bachman, Hjelm, and Buchwalter 2019; Chen et al. 2020; He et al. 2020b; Henaff 2020; Hjelm et al. 2018; Misra and Maaten 2020; Oord, Li, and Vinyals 2018; Wu et al. 2018) and get the state of the art performance. The learned image features could be utilized for downstream tasks such as image retrieval or fine-tuning for classification. In our work, we take advantage of the SimCLR (Chen et al. 2020) framework and fine-tune the learned features with psychometric testing for semantic image representations.

## Semantic-Based Few-Shot Learning

Our proposed framework contains three parts. First, as we aim to tackle the limitation caused by label-based supervision, we assume no label information is provided in advance. Self-supervised learning (SSL) is applied for representation learning in the first stage. Next, we adopt a psychometric testing procedure (Gescheider 2013) that relies on discriminative testing to obtain transferable semantic conception relationships. The elicited conception similarities are then used to fine-tune the features learned by SSL using a multi-layer perceptron (MLP) (Friedman 2017) in a semantic representation network. In the last stage, with the fine-tuned network, we could search for each query’s most semantically similar image in support set by Euclidean distances, even when the target and query images are not sharing the same class. We illustrate our whole framework in Figure 2.

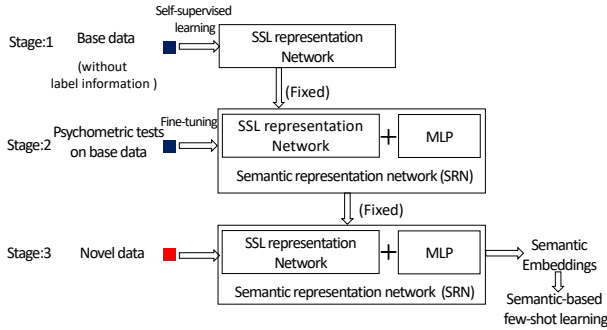


Figure 2: Overview of the proposed method.

### Problem Formulation

Consider the situation we are given a base dataset contains classes  $C_{base}$  with adequate labeled images, and a novel dataset contains classes  $C_{novel}$  where only a few labeled samples are available per class. There is no overlapping with these two datasets, i.e.,  $C_{base} \cap C_{novel} = \emptyset$ . The general idea of the few-shot problem is taking advantage of the sufficient labeled samples in  $C_{base}$  to obtain a good classifier for the novel class  $C_{novel}$ . In a standard  $N$ -way  $K$ -shot classification task, we random sample  $N$  classes from novel class  $C_{novel}$  with  $K$  samples per class to form the support set, and sample query images from the same  $N$  classes to create the query set. We aim to classify the query images into these  $N$  classes based on the support set.

Then we extend the problem to the semantic-based few-shot learning scenario. Assume we have a conception tree  $G = (V, E)$  where  $V$  means the nodes and  $E$  are edges. The bottom layer class  $C = c_1, \dots, c_n \in V$  denotes the lowest level of concepts that we concern, and could merge to more general concepts (superclass nodes) if they are conceptually similar. An example for such a structure is given in Figure 3. The base class  $C_{base}$  and novel class  $C_{novel}$  are represented as the leaf nodes and share the same superclasses nodes. As we aim to solve this problem without label-based supervision, we are not able to specify a few-shot task using the label information as the typical few-shot learning setting, i.e., sampling multiple images with the same labels to create a

class in support set. Therefore, we random sample  $N$  image without specifying their classes from the  $C_{novel}$  to build the support set and sample one image as a query to form a  $N$ -way  $1$ -shot semantic-based few-shot learning task. Our goal is to find the most semantically similar image from support set to a query.

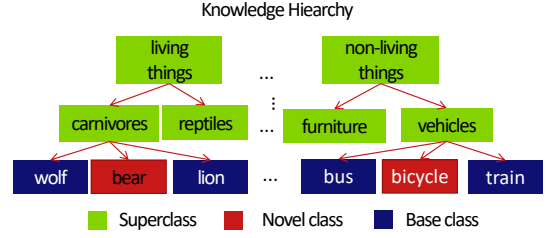


Figure 3: CIFAR-100 with knowledge hierarchy.

The semantic distance between two samples  $(x, y)$  is defined by the height of the lowest common subsumer (LCS) of these samples divided by the height of the hierarchy (Barz and Denzler 2019; Verma et al. 2012):

$$D_s(x, y) = \frac{\text{height}(\text{lcs}(x, y))}{\max_{w \in V} \text{height}(w)} \quad (1)$$

As  $D_s(x, y)$  ranges from 0 to 1, we could define the semantic similarity by:

$$S_s(x, y) = 1 - D_s(x, y) \quad (2)$$

An example could be seen in Figure 3. The LCS of *wolf* and *lion* are *carnivores*, and the height of the hierarchical tree is 3. Therefore  $D_s(\text{wolf}, \text{lion}) = \frac{1}{3}$ , and  $S_s(\text{wolf}, \text{lion}) = \frac{2}{3}$ . Note that the typical few-shot learning is a special case when  $S_s(x, y) = 1$ , in which  $x$  is the query image,  $y$  is from support set, and  $x, y$  belong to a same leaf node.

### Self-Supervised Feature Learning

We use self-supervised learning to learn the image features from  $C_{base}$  before using psychometric testing for fine-tuning. SimCLR (Chen et al. 2020) framework is applied in our work for its conciseness and good performance. It learns representation by maximizing the similarity between two views (augmentations) of the same image.

From  $C_{base}$ , we randomly sample  $N$  images each batch and create two random augmentation views for each image to form  $2N$  data points. Each data pair generated from the same image is considered a positive pair, or a negative pair if it’s from different images. The contrastive loss function for a mini-batch could be written as:

$$L_{self} = - \sum_{i=1}^{2N} \log \frac{\exp(\text{sim}(z_i, z_{j(i)})/\tau)}{\sum_{a \in A(i)} \exp(\text{sim}(z_i, z_a)/\tau)} \quad (3)$$

where  $z_i = g(f(x_i))$ ,  $f(\cdot)$  a neural network called encoder to extract features from augmented images,  $g(\cdot)$  is the projection head that maps features to a space where contrastive loss

is applied. Cosine similarity  $\text{sim}(u, v)$  is adopted to measure the similarity of  $u$  and  $v$  by the dot product between their  $L_2$  normalized features.  $\tau$  denotes a scalar temperature parameter.  $i$  is the index or all the  $2N$  augmented views of images.  $j(i)$  is the index of positive view to image  $i$  and  $A(i)$  is the set of all indices except  $i$ .

### Psychometric Testing

Different from label-based supervision, we apply three-alternative-force choice (3AFC) (DeCarlo 2012) psychometric tests to elicit the semantic perceptions from  $C_{base}$ . These perceptions could be transferred to  $C_{novel}$  through the shared high-level conceptions (superclasses) in the hierarchical knowledge tree (as shown in Figure 3).

To be specific, we sample three images from  $C_{base}$  and ask the annotators to choose the most dissimilar one (see Figure 4). By carrying this simple task, perceptions of conception similarities are obtained.

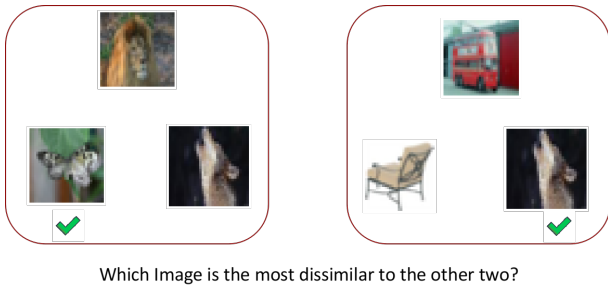


Figure 4: Examples of two 3AFC psychometric testings. In the first test, annotators tend to choose the butterfly as the most dissimilar one since other two are large carnivores. In the second test, annotators are more likely to choose the wolf because it is the only living things.

Next, a semantic representation network (SRN) is built to map these perceived conception similarities to embedding distances. Specifically, we add a multi-layer perceptron (MLP) with a single hidden layer on top of the representations learned from SSL, freeze the SSL network, and fine-tune the MLP by the following dual-triplet loss function (Yin, Menkovski, and Pechenizkiy 2020; Yin et al. 2021):

$$L = \sum_{i=1}^N [d(x_{p1}^i, x_{p2}^i) - d(x_n^i, x_{p1}^i), +m]_+ + [d(x_{p1}^i, x_{p2}^i) - d(x_n^i, x_{p2}^i), +m]_+ \quad (4)$$

where  $x_n$  is the negative image chosen by annotator,  $x_{p1}, x_{p2}$  are two unpicked positive images that have closer concept similarity at the 3AFC tests (see Figure 4).  $d(x, y)$  denotes the these two points' Euclidean distance between the normalized features extracted by our semantic representation network.  $N$  is the number of psychometric tests in a mini-batch.

This loss function encourages images that the annotator perceives similar to be close to each other and enforces a distance margin  $m$  between positive pairs and negative pairs.

### Semantic-Based Few-Shot Prediction

After fine-tuning our proposed network with 3-AFC tests from  $C_{base}$ , we could extract visual features for image samples from  $C_{novel}$  using this network and apply the nearest neighbor search method for semantic-based few-shot learning prediction. Specifically, for a query image in a task, we compute its normalized Euclidean distance to each support sample and find the nearest one, which is the predicted most semantically similar image to the query when considering a higher-level concept.

### Experiments and Discussion

Since the label-based supervision is a bottleneck that limits the models' potential in the semantic-based few-shot learning setting, we assume no label information and no conception structure are preprovided for both  $C_{base}$  and  $C_{novel}$ . However, we are then not able to assess whether the semantic assignment to query is correct using the defined semantic similarity metric (see Equation 2). Therefore, we simulate a virtual annotator who always precisely responds to the 3AFC tests based on a given knowledge hierarchy, so that the accuracy could be measured in an objective manner by this semantic similarity.

We evaluate our model on CIFAR-100 dataset under three metrics: the typical few-shot learning accuracy, the semantic-based few-shot learning accuracy, and the required annotation numbers. Then we investigate how the number of psychometric test responses impacts the model's performance. Besides, a TSNE visualization (Van der Maaten and Hinton 2008) of the learned features is plotted for an intuitive understanding.

**Dataset** We use the CIFAR-100 in our experiment and build an inner conception hierarchy tree based on the preprovided coarse and finer labels. Besides, we build another layer on top of the coarse level labels by distinguishing living from non-living things. A three-layer conception tree is then created, which includes 2, 10, 100 nodes from top to bottom layers, as illustrated in Figure 3. 60 classes are randomly sampled from the bottom layer as base classes, and the rest 40 classes are used for novel classes.

**Few-shot learning accuracy** Note when there is a label matching between the query image and support images, i.e., the semantic similarity is equal to 1, the semantic-based few-shot learning problem is then transmitted to a typical few-shot learning problem. We choose the prototypical network (Snell, Swersky, and Zemel 2017) as a baseline and compared it with our proposed method in both typical few-shot learning accuracy and semantic-based few-shot learning accuracy.

In our work, we use the SGD optimizer with momentum 0.9, and set the decay factor to 0.1. When extracting image features in SSL, ResNet50 (He et al. 2016) is applied as backbone and are trained for 1000 epochs with 128 batch-size. The learning rate decays from 0.5 at epoch 700, 800 and 900. When fine-tuning by psychometric responses, margin value, learning rate, training epochs are set to 0.4, 0.001, 15 respectively. During prototypical network training, we use the same backbone of SSL for a fair comparison, train



Table 1: Comparison with the baseline.

Model	Annotation Type	Number of Annotations ( $C_{base}$ )	5-way 1-shot Acc(%)		20-way 1-shot Acc(%)	
			Typical	Semantic	Typical	Semantic
PN (Snell, Swersky, and Zemel 2017)	Label Based	36000	<b>57.52</b>	42.37	<b>31.18</b>	19.81
SRN(Ours)	Psychometric Testing	<b>1000</b>	52.57	<b>52.35</b>	28.75	<b>27.16</b>

the model 100 epochs with 10000 tasks each epoch, and set the learning rate to 0.1 that decays every 20 epochs.

The results are reported in Table 1. It could be seen without losing too much accuracy of typical one-shot learning (decreasing by 4.95% in 5-way, and 2.43% in 20-way). We could boost the ability of semantic-based one-shot learning significantly (increasing by 9.98% in 5-way, and 7.35% in 20-way). Furthermore, the annotation burdens on base data are dramatically released from 36000 times label-based annotations to 1000 times psychometric testings.

**Impact of the number of psychometric test responses** We train our model using 500 psychometric tests in the first iteration and add 500 more tests to retrain the model in each of the following iterations. The model is evaluated under the 5-way 1-shot scenario and we plot the results in Figure 5. It could be noticed that the accuracy of typical few-shot learning remains steady with different numbers of psychometric tests. That is because our psychometric tests only aim to provide semantic constrain rather than learning discriminative features. We also find that the ability of semantic few-shot learning gets a noticeable improvement when increasing training samples from 500 to 1000 tests but keeps stuck after that. The possible reason might be that with the help of pre-trained SSL features, we could easily get a high accuracy using only a few psychometric tests. However, as we only fine-tune on MLP without training the whole network, the semantic few-shot accuracy would quickly reach a bottleneck even with more psychometric responses.

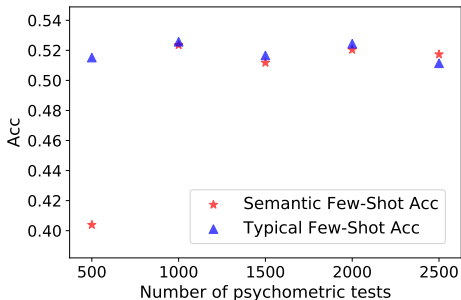


Figure 5: 5-way 1-shot learning accuracy under different number of psychometric tests.

**TSNE visualization** We visualize the embedding features of five categories randomly chosen from  $C_{novel}$  (See Figure 6). It could be seen that with our proposed method, categories that are similar in concepts tend to be closer to each other. For example, all the non-living things (mountain, forest, streetcar) are located in the top area while living things (bee, tiger) are placed bottom. Mountain and forest are the nearest

two clusters since they are “all outdoor scenes” and their semantic distance is the closest among the five categories. On the other hand, the prototypical network could successfully separate the five categories apart from each other, but they are located randomly in the graph without considering the semantic relationships.

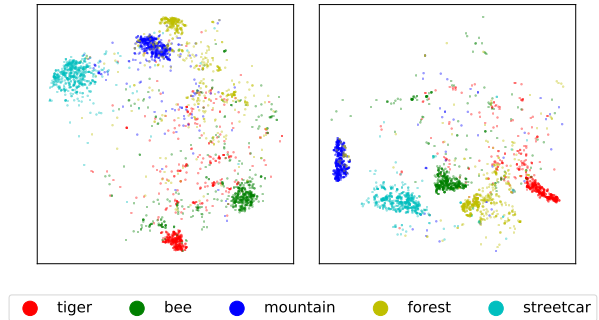


Figure 6: Embedding visualization after TSNE with our proposed method (left), and prototypical network (right).

## Conclusion

Few-shot learning is typically under label-based supervision, which discards the semantic relationships and fails to make a class association when there is no label matching between support and query set. However, humans could easily identify the right association by considering a higher-level concept. Inspired by this, we present a psychometric testing based method that could capture images’ high-level conception relationships to address the challenge. We evaluate our method on CIFAR-100 dataset. The results indicate that our method is capable of achieving higher semantic-based few-shot learning accuracy even with fewer annotating burdens than the baseline.

## References

- Bachman, P.; Hjelm, R. D.; and Buchwalter, W. 2019. Learning representations by maximizing mutual information across views. *arXiv preprint arXiv:1906.00910*.
- Barz, B.; and Denzler, J. 2019. Hierarchy-based image embeddings for semantic image retrieval. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 638–647. IEEE.
- Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, 1597–1607. PMLR.

- DeCarlo, L. T. 2012. On a signal detection approach to m-alternative forced choice with bias, with maximum likelihood and Bayesian approaches to estimation. *Journal of Mathematical Psychology*, 56(3): 196–207.
- Fechner, G. T. 1860. *Elemente der psychophysik*, volume 2. Breitkopf u. Härtel.
- Feng, H.-C.; Marcellin, M. W.; and Bilgin, A. 2014. A methodology for visually lossless JPEG2000 compression of monochrome stereo images. *IEEE Transactions on Image Processing*, 24(2): 560–572.
- Finn, C.; Abbeel, P.; and Levine, S. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, 1126–1135. PMLR.
- Friedman, J. H. 2017. *The elements of statistical learning: Data mining, inference, and prediction*. Springer open.
- Gescheider, G. A. 2013. *Psychophysics: the fundamentals*. Psychology Press.
- Gidaris, S.; and Komodakis, N. 2018. Dynamic few-shot visual learning without forgetting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4367–4375.
- Gidaris, S.; Singh, P.; and Komodakis, N. 2018. Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728*.
- He, J.; Hong, R.; Liu, X.; Xu, M.; Zha, Z.-J.; and Wang, M. 2020a. Memory-augmented relation network for few-shot learning. In *Proceedings of the 28th ACM International Conference on Multimedia*, 1236–1244.
- He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020b. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9729–9738.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Henaff, O. 2020. Data-efficient image recognition with contrastive predictive coding. In *International Conference on Machine Learning*, 4182–4192. PMLR.
- Hjelm, R. D.; Fedorov, A.; Lavoie-Marchildon, S.; Grewal, K.; Bachman, P.; Trischler, A.; and Bengio, Y. 2018. Learning deep representations by mutual information estimation and maximization. *arXiv preprint arXiv:1808.06670*.
- Krizhevsky, A.; Hinton, G.; et al. 2009. Learning multiple layers of features from tiny images.
- Lake, B.; Salakhutdinov, R.; Gross, J.; and Tenenbaum, J. 2011. One shot learning of simple visual concepts. In *Proceedings of the annual meeting of the cognitive science society*, volume 33.
- Larsson, G.; Maire, M.; and Shakhnarovich, G. 2016. Learning representations for automatic colorization. In *European conference on computer vision*, 577–593. Springer.
- Misra, I.; and Maaten, L. v. d. 2020. Self-supervised learning of pretext-invariant representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6707–6717.
- Nichol, A.; and Schulman, J. 2018. Reptile: a scalable meta-learning algorithm. *arXiv preprint arXiv:1803.02999*, 2(3): 4.
- Noroozi, M.; Pirsiavash, H.; and Favaro, P. 2017. Representation learning by learning to count. In *Proceedings of the IEEE International Conference on Computer Vision*, 5898–5906.
- Oord, A. v. d.; Li, Y.; and Vinyals, O. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.
- Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; and Efros, A. A. 2016. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2536–2544.
- Snell, J.; Swersky, K.; and Zemel, R. S. 2017. Prototypical networks for few-shot learning. *arXiv preprint arXiv:1703.05175*.
- Son, I.; Winslow, M.; Yazici, B.; and Xu, X. 2006. X-ray imaging optimization using virtual phantoms and computerized observer modelling. *Physics in Medicine & Biology*, 51(17): 4289.
- Sung, F.; Yang, Y.; Zhang, L.; Xiang, T.; Torr, P. H.; and Hospedales, T. M. 2018. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1199–1208.
- Trinh, T. H.; Luong, M.-T.; and Le, Q. V. 2019. Selfie: Self-supervised pretraining for image embedding. *arXiv preprint arXiv:1906.02940*.
- Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).
- Verma, N.; Mahajan, D.; Sellamanickam, S.; and Nair, V. 2012. Learning hierarchical similarity metrics. In *2012 IEEE conference on computer vision and pattern recognition*, 2280–2287. IEEE.
- Vinyals, O.; Blundell, C.; Lillicrap, T.; Wierstra, D.; et al. 2016. Matching networks for one shot learning. *Advances in neural information processing systems*, 29: 3630–3638.
- Wu, Z.; Xiong, Y.; Yu, S. X.; and Lin, D. 2018. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3733–3742.
- Yin, L.; Menkovski, V.; Liu, S.; and Pechenizkiy, M. 2021. Hierarchical Semantic Segmentation using Psychometric Learning. *arXiv preprint arXiv:2107.03212*.
- Yin, L.; Menkovski, V.; and Pechenizkiy, M. 2020. Knowledge Elicitation using Deep Metric Learning and Psychometric Testing. *arXiv preprint arXiv:2004.06353*.
- Zhang, R.; Isola, P.; and Efros, A. A. 2016. Colorful image colorization. In *European conference on computer vision*, 649–666. Springer.
- Zhang, R.; Isola, P.; and Efros, A. A. 2017. Split-brain autoencoders: Unsupervised learning by cross-channel prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1058–1067.