One-Shot Image Recognition Using Prototypical Encoders with Reduced Hubness

Chenxi Xiao
Purdue University
West Lafayette, 47906, USA
xiao237@purdue.edu

Naveen Madapana
Purdue University
West Lafayette, 47906, USA
nmadapan@purdue.edu

Juan Wachs
Purdue University
West Lafayette, 47906, USA
jpwachs@purdue.edu

Abstract

Humans have the innate ability to recognize new objects just by looking at sketches of them (also referred as to prototype images). Similarly, prototypical images can be used as an effective visual representations of unseen classes to tackle few-shot learning (FSL) tasks. Our main goal is to recognize unseen hand signs (gestures) traffic-signs, and corporate-logos, by having their iconographic images or prototypes. Previous works proposed to utilize variational prototypical-encoders (VPE) to address FSL problems. While VPE learns an image-to-image translation task efficiently, we discovered that its performance is significantly hampered by the so-called hubness problem and it fails to regulate the representations in the latent space. Hence, we propose a new model (VPE++) that inherently reduces hubness and incorporates contrastive and multi-task losses to increase the discriminative ability of FSL models. Results show that the VPE++ approach can generalize better to the unseen classes and can achieve superior accuracies on logos, traffic signs, and hand gestures datasets as compared to the state-of-the-art.

1. Introduction

Machine learning algorithms generally require large amounts of data to capture the underlying data distribution and to effectively perform classification tasks. Conversely, humans have the innate ability to generalize across new categories or unknown scenarios by efficiently utilizing their past experience [58]. Hence, there is a significant semantic gap in the way humans and machines learn to recognize objects. This has led to the advent of modern learning paradigms such as *few-shot* learning (FSL), which aims to recognize the unseen categories by having only a few observations [9].

In this context, *k-shot* refers to the problem in which there are only k samples for each unseen category. In the extreme case, where the value of $k=1$, the model is expected to recognize a new category by just having one labeled example.

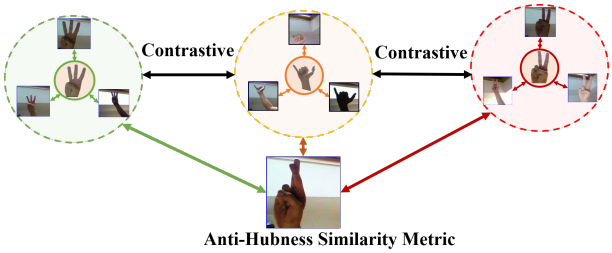


Figure 1. Illustration of spatial relationship between real images and their prototypes (image in the center). We utilize contrastive learning to enhance the mutual information between prototypes and real images in the latent space.

Note that, the quality of the sample available for re-training greatly affects the performance of the system. In this paper, we focus on graphical symbols, sketches, or icons that compactly convey the semantic meaning of the categories, as they can be used as a single available training example to tackle the problem of *one-shot* learning. Moreover, such symbols, referred to as *prototypes* (refer to the clustering centroid in Fig. 1), contain rich contextual and semantic information embedded in them which are proven to be helpful in the *one-shot* learning tasks [22].

While humans can easily comprehend such prototypical abstractions, machines face several challenges due to the inherent perceptual gap between the prototypes and the real images [22, 27]. In other words, the real images undergo perspective transformations, distortions, variable lighting conditions and occlusions, which makes data distribution of the real images and their prototypes considerably different from each other. Due to these challenges, very few research works have successfully applied prototypical images in FSL tasks. For instance, Kim *et al*. proposed Variational Prototyping-Encoder (VPE) that learns a functional mapping that translates the real images into their prototypes [27]. While their method demonstrated a feasible way of utilizing prototypical images, we noticed two major limitations of VPE that hamper its performance on FSL.

First, VPE suffers from the data hubness problem i.e. a

small subset of unseen samples tends to act as hubs and they are more frequently retrieved than other samples. In other words, some test samples are rarely retrieved even when the similarity with the query image is high (refer to Sections 2.3 and 3.2). Second, VPE does not impose any additional constraints in the latent space i.e. it does not explicitly force the latent representation of a real image and its corresponding prototype to be equivalent. Hence, the data points in the high-dimensional latent space that belong to multiple classes can still distribute in ways that are difficult to discriminate (refer to Sec. 2.4 and Sec. 3.3).

In this regard, we propose a new model (referred to as VPE++) to address these limitations. Kim et al. tested the VPE approach on logos and traffic sign datasets. We conducted experiments on these datasets to show that VPE++ outperforms VPE by a significant margin. Furthermore, we demonstrated the generalizability of our approach using the American Sign Language (ASL) hand gesture dataset. The rationale behind testing VPE++ on hand signs lies in the fact that a significant part of the ASL community learns those gestures from sketches, icons, and pictorial representations that are complementary to face to face demonstration [6]. In fact, ASL allows for an integration of visual imagery with the linguistic structure on a magnitude not seen in other spoken languages [51]. The same analogy can be made about driving lessons - traffic signals are often learned from iconic representations in instruction booklets as a pre-requisite for passing the driver license's test [23].

The main contributions of this work are as follows. First, we demonstrate through experiments that the hubness issue is prevalent in the latent space of VPE and thereby, in the FSL problems. Next, we propose to utilize CL2N similarity metric to address the issue of data hubness and show that it can compensate for cases of unbalance that individual samples being retrieved during the testing stage. Second, we incorporate contrastive loss which allows better inter-class discriminability when using prototypes. Lastly, we integrate multi-task learning and deep supervision into our model to improve the gradient backpropagation and model generalization. Our code is released at: `github.com/MegaYEye/VPEReducedHubness.git`

2. Related Work

2.1. Few-shot Learning

Conventional machine learning methods work well under the assumption that training and testing data belong to the same distribution. In contrast, *transfer learning* allows the domains, tasks and data distribution of training and test data to be significantly different [41, 32]. The task of *few-shot* learning (a sub-paradigm of transfer learning) was first introduced by Fei et al. [9] and was extensively studied in domains related to *object* classification [10, 50, 29], *face* recognition [45, 17], and *action* recognition [36, 2, 16]. For example, Florian et al. proposed an approach to learn unified embeddings of faces to improve real-time face recognition [45].

Seminal works in *few-shot* learning include but not limited to [9, 29, 35, 13, 38, 56]. Works such as [56, 47, 50] extracted and transformed the task-related information to a metric space, in which the classification was performed through comparing the similarity scores. These approaches fall under the category of metric learning. Conversely, [35, 13, 38] focus on endowing the ability to adapt to new tasks, and thereby belong to the category of meta-learning.

2.2. Few-Shot Learning Based on Prototypes

In this work, our primary objective is to perform *one-shot* recognition with prototypical images as the only available examples for the new classes. The intuition behind utilizing prototypes lies in the fact that they carry rich contextual information that can be leveraged upon for recognizing new classes [27]. Several attempts were made in the past to effectively utilize the prototypes as a part of the one-shot framework. For instance, [47, 56, 33] constructed "prototypes" from statistical information obtained from the training samples.

In this regard, a prototype would act as a cluster center for samples belonging to a particular category. However, the estimation bias of the clustering center can increase drastically when the data distribution in the latent space is unbalanced, or the data is relatively sparse. Furthermore, [26, 27] proposed a novel framework in which variational auto-encoders were adapted to take real images as the input and reconstruct their prototypes instead of reconstructing the inputs. Their framework does not consider prototypes as outliers and significantly reduces the estimation bias of the latent representations.

2.3. Hubness Reduction

Hubness is a phenomenon in which a small subset of samples act as *hubs* or *universal neighbors* and are more frequently retrieved by neighborhood searching algorithms than other samples [7, 42]. In other words, a small set of samples are observed too frequently while other samples are rarely retrieved. As an intrinsic problem in high-dimensional space, the hubness issue is known to have negative effects on many representation learning tasks such as word translation [5, 25, 15] and image retrieval [7]. Although the hubness has been studied in detail in natural language processing tasks, it has been hardly analyzed in the context of FSL.

Several hubness-aware machine learning techniques were developed in the past in order to alleviate such adverse effects. For instance, [25] proposed an approach that explicitly penalizes the hubness-aware loss function known as RCSLS. On the other hand, some works passively reduced the hubness on the embedding samples in the latent space.

Such approaches generally rely on improved neighbourhood searching strategies that can increase the chance of discovering samples that are being visited with low possibility. These approaches include NHBNN [54], HWKNN [53], HFNN [55], CSLS [5], and ISF [46]. However, these methods are only available for embedding spaces with abundant observations (not suitable for few-shot learning).

In this paper, we discover that the hubness has a negative effect on the accuracy of the metric-based few-shot learning tasks. The FSL task differs from the aforementioned problems in the number of observations of the target domain that are available. Since FSL problems have only a few examples from the target domain, calculating a hubness score for each individual sample is non-trivial. We address this issue by utilizing a recently proposed neighborhood searching approach known as CL2N [57].

2.4. Contrastive Loss and Auxiliary Supervision

Contrastive learning aims to learn representations so that samples corresponding to the same category (or instance) are pushed together and the samples belonging to distinct categories are pulled away from each other [33]. Contrastive learning is known to have the ability to improve generalization capacity [3, 33, 18, 39, 52]. One main reason for this is that overfitting can be prevented by increasing the class margin [59]. It was shown that contrastive learning coupled with self-supervised learning can achieve performances comparable with many supervised learning approaches [3, 18].

Recently, machine learning approaches began to leverage on **auxillary supervision** to improve performance on multiple tasks and thereby, improve generalization. For classification tasks, deep supervision [30] induces side-branches to accelerate convergence and alleviate the vanishing gradients problem. In [31], a step ahead is taken by enhancing the gradient agreement from different branches. Their experiments show that deep supervision can improve the prediction accuracy and generalization.

Another mainstream method is **multi-task learning**, where supervision is accomplished using different tasks. The shared parameters learned across different tasks can better capture the underlying data distribution and improve generalization [14]. In this work, we integrate both deep supervision and multi-task learning to improve model generalization.

3. VPE++: Improved Prototypical Encoder

Our approach to few-shot learning from prototypes builds on the variational prototypical-encoder (VPE) model [27] by addressing its key limitations. Our model, referred to as VPE++, leverages on the VPE by jointly optimizing multi-task classification and contrastive latent losses in addition to prototypical reconstruction loss. In this section, we discuss the issue of hubness in the context of FSL and present three major components of VPE++ model which led to significant improvements in the performance on FSL tasks.

3.1. VPE Backbone

Given a real image $\boldsymbol{x}$, VPE [27] learns a mapping that transforms the input image $\boldsymbol{x}$ to its corresponding prototype $\boldsymbol{t}$. This is achieved through an auto-encoder paradigm in which the *encoder* maps the raw input image $\boldsymbol{x}$ from image space to an embedding $\boldsymbol{z_x}$ in the latent space while the *decoder* maps these latent embeddings to the corresponding prototypes $\boldsymbol{t}$ in the image space. Kim et al. showed that the generalization of VPE can be significantly improved by augmenting the training data $(\boldsymbol{x}, \boldsymbol{t})$ via random rotations and flips, and by utilizing spatial transformations [21] through the encoder to counteract the variances of inputs.

VPE++ model utilizes the original network structure of VPE to construct its encoder and decoder components. However, we recognize that utilizing the VPE architecture directly will result in a hubbed latent representation that reduces inter-class discriminability. Our key improvements lie in model optimization procedures that involve additional training objectives. For instance, the parameters of the VPE model i.e. the parameters of the encoder $p_\theta(\boldsymbol{z_x}|\boldsymbol{x})$ and the decoder $p_\theta(\boldsymbol{t}|\boldsymbol{z_x})$ are learned by minimizing the prototypical reconstruction loss given in Eq. (2), which does not fully exploit the prior knowledge from the templates. In this context, VPE++ model tackles this issue by simultaneously optimizing the prototypical latent/contrastive loss $\mathcal{L}_{PCCL}$ and multi-task label prediction loss $\mathcal{L}_{mtl}$ in addition to the reconstruction loss. The overall loss function of the VPE++ is a weighted sum of these three components (refer to Eq. (1)). An overview of our method during training is shown in Fig. 2.

$$\mathcal{L}_{VPE++} = \mathcal{L}_{recon} + \alpha\mathcal{L}_{PCCL} + \beta\mathcal{L}_{mtl} \tag{1}$$

3.2. Hubness Reduction

Hubness is a recurrent problem that is reportedly present in the latent space i.e. the latent vectors that are closer to the centroid are more likely to act as hubs [61]. In this section, we show that the VPE suffers from the *so-called* hubness problem and propose to utilize CL2N metric to effectively tackle this issue.

To prove the existence of the hubness problem, we visualized the latent space of the VPE model by transforming it into two-dimensional t-SNE embeddings, as shown in Fig. 3 (a). Overall, 9585 samples obtained from the Flickr32 dataset (refer to Sec. 4) were visualized. For each sample, we retrieved its nearest 15 neighbouring samples based on Euclidean distance in the embedding space, and then used color (blue and red color indicate lowest and highest frequencies respectively) to indicate how many times a sample was retrieved in the t-SNE visualization. Furthermore, a

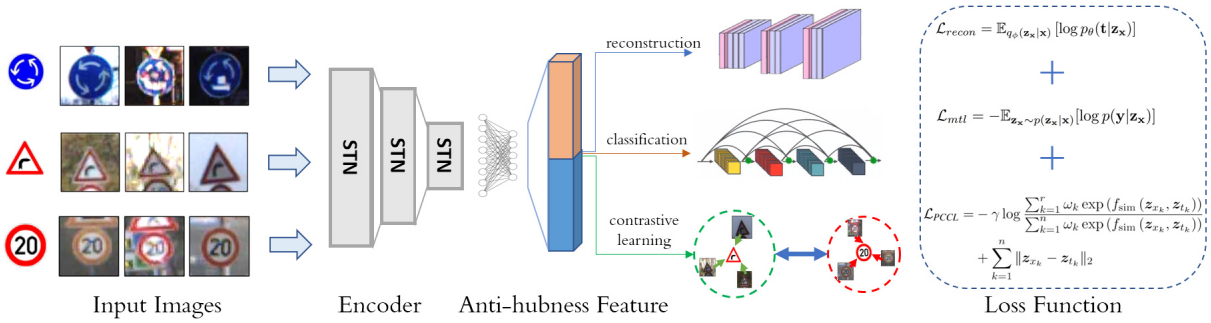
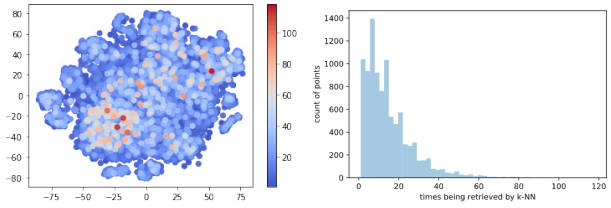


Figure 2. Training pipeline of the VPE++ model. The VPE++ architecture utilizes Spatial Transformer Network (STN) for feature encoding and simultaneously alleviates the issues related to *data hubness* and *vanishing gradients* by minimizing a combined loss function $L_{VPE++} = \mathcal{L}_{recon} + \alpha\mathcal{L}_{PCCL} + \beta\mathcal{L}_{mtl}$.

histogram of the retrieval count of samples was shown in Fig. 3 (b).

It can be noticed that Fig. 3 (b) resembles a long tail distribution i.e. the histogram is not centered at the average value (15) and was skewed towards the left side. In other words, a small subset of samples was visited too frequently (hubs) while a considerable number of samples were rarely visited. Hence, there were few red circles and a large number of blue/white circles in Fig. 3 (a). This experiment shows that the VPE suffers from the *hubness problem* i.e. the similarity comparison in the latent space has an inherent bias towards a small set of frequently visited samples.



(a) (b)

Figure 3. (a) Visualization of the latent space using t-SNE embeddings. We performed k-NN with $k = 15$ on all samples, and use color to indicate how many times a sample has been retrieved. (b) The corresponding histogram plot showing that a small subset of samples (hubs) was visited too frequently while a considerable number of samples were rarely visited.

Our approach to tackle the hubness issue in few-shot learning is two-fold. The *first improvement* lies in altering the VPE loss function to reduce hubness to an extent. VPE suffers from the hubness problem partially due to the assumption that the latent space is a Gaussian sphere i.e. embeddings follow a Gaussian distribution. This can be shown by investigating the VPE loss function (refer to Eq. (2)).

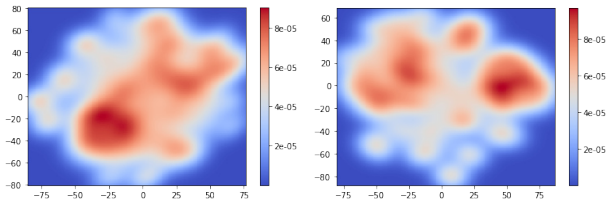
$$\begin{aligned}\log p_\theta(\mathbf{t}) =& \mathbb{E}_{q_\phi(\mathbf{z_x}|\mathbf{x})}\left[\log p_\theta(\mathbf{t}|\mathbf{z_x})\right] \\ &- D_{KL}\left[q_\phi(\mathbf{z_x}|\mathbf{x})\|p_\theta(\mathbf{z_x})\right]\end{aligned} \tag{2}$$

The first loss term $\mathbb{E}_{q_\phi(\mathbf{z_x}|\mathbf{x})}\left[\log p_\theta(\mathbf{t}|\mathbf{z_x})\right]$ is used to maximize the likelihood of reconstructing prototype $\mathbf{t}$. The second loss term $D_{KL}\left[q_\phi(\mathbf{z_x}|\mathbf{x})\|p_\theta(\mathbf{z_x})\right]$ is the KL divergence between latent data distribution $p(\mathbf{z_x}|x)$ and the distribution prior $p_\theta(\mathbf{z_x})$. Given the intractability of obtaining the true latent distribution, VPE assumes that $p_\theta(\mathbf{z_x})$ follows a Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$, where $\boldsymbol{I}$ is an identity matrix. This Gaussian prior is induced by the variational approximation and is known to improve generalization due to the inherent regularization effect [44, 60]. On the contrary, the recent studies in contrastive learning suggest that enlarging the inter-class distance can be more beneficial in achieving better generalization abilities as opposed to shrinking the latent space to a Gaussian sphere [3, 40]. Furthermore, the probability density function associated with $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$ is known to aggravate the issue of hubness as the samples are more likely to distribute around the origin, thus having higher likelihood of acting as hubs [4, 61]. Lastly, while the Gaussian prior enables sampling in the latent space by modeling it as a continuous and a disentangled distribution [28, 19], our approach does not rely on this effect as we do not sample from the latent space. Therefore, we believe that there is no need to induce the Gaussian prior in our methodology. Further, we show in our experiments that assuming the latent space to follow a Gaussian distribution has a negative effect on performance. That is, optimizing solely on the loss function given in Eq. (3) can improve performance considerably in comparison to VPE.

$$\mathcal{L}_{recon} = \mathbb{E}_{q_\phi(\mathbf{z_x}|\mathbf{x})}\left[\log p_\theta(\mathbf{t}|\mathbf{z_x})\right] \tag{3}$$

A t-SNE visualization of the embedding space generated by optimizing the Eq. (2) and Eq. (3) using Flickr32 dataset is shown in the Fig. 4 (a) and 4 (b), respectively. Intuitively, Fig. 4 (b) has distinct boundaries between the clusters, and the density around origin is smaller compared to Fig. 4 (a). This shows that the issue of hubness is reduced and the discriminability between classes is improved.

The *second improvement* for anti-hubness is to incorporate a hubness reduction scheme during retrieval. Hence, we experimented with several hubness reduction schemes and found the following three strategies to be equally effective in our one-shot tasks: 1. Nearest Neighbourhood with Centered and L2-normalized feature (CL2N) [57], 2. K-Nearest



(a) (b)

Figure 4. Visualization of the latent space using t-SNE embeddings (color indicates the sample density): (a) The embedding space distribution obtained by optimizing the original VPE loss function (Eq. (2)). (b) The embedding space distribution obtained by optimizing the improved loss (Eq. (3))

Neighbor Graph (K-NNG) [8, 12], and 3. Hierarchical Navigable Small World (HNSW) [34, 12]. However, we used CL2N as the default metric for our algorithm due to its low computational cost. The procedure of incorporating CL2N with our prototype based 1-shot learning paradigm is given as follows. First, we calculated the mean latent embeddings of all prototypes $\boldsymbol{z_t}$ (constructed by the encoder), and subtract it from the embeddings of candidate template image $\boldsymbol{z_t}$ and the query image $\boldsymbol{z_x}$, as shown in Eq. (4).

$$\begin{cases} \hat{\boldsymbol{z}}_{\boldsymbol{t}} = \boldsymbol{z_t} - \frac{1}{K}\sum_{t\in\mathcal{D}_{\text{test}}} \boldsymbol{z_t} \\ \hat{\boldsymbol{z}}_{\boldsymbol{x}} = \boldsymbol{z_x} - \frac{1}{K}\sum_{t\in\mathcal{D}_{\text{test}}} \boldsymbol{z_t}, \end{cases} \tag{4}$$

Next, we performed L_2-normalization on both the $\hat{\boldsymbol{z}}_{\boldsymbol{t}}$ and $\hat{\boldsymbol{z}}_{\boldsymbol{x}}$, as shown in Eq. (5). This step reduces the chance of being retrieved for samples located in dense areas [5]. One-shot learning is then formulated as a retrieval task by k-NN with L_2 metric on the $\hat{\boldsymbol{z}}_{\boldsymbol{t}}$ embedding set queried by $\hat{\boldsymbol{z}}_{\boldsymbol{x}}$.

$$\begin{cases} \hat{\boldsymbol{z}}_{\boldsymbol{t}} \leftarrow \frac{\hat{\boldsymbol{z}}_{\boldsymbol{t}}}{||\hat{\boldsymbol{z}}_{\boldsymbol{t}}||_2} \\ \hat{\boldsymbol{z}}_{\boldsymbol{x}} \leftarrow \frac{\hat{\boldsymbol{z}}_{\boldsymbol{x}}}{||\hat{\boldsymbol{z}}_{\boldsymbol{x}}||_2} \end{cases} \tag{5}$$

3.3. Prototype Centered Contrastive Loss (PCCL)

Contrastive learning has been widely shown to be capable of improving the model generalization [3, 33, 45]. Hence, we developed a novel loss referred to as Prototype Centered Contrastive Loss (PCCL) that facilitates contrastive learning to be applied for the prototypical images.

Given a set of samples (i.e. a minibatch) in the latent space $\mathcal{Z} = [\boldsymbol{z_{x_1}}, \boldsymbol{z_{x_2}}, ..., \boldsymbol{z_{x_r}}, ..., \boldsymbol{z_{x_n}}]$, assume that a subset of them $\mathcal{Z}_s = [\boldsymbol{z_{x_1}}, \boldsymbol{z_{x_2}}, ..., \boldsymbol{z_{x_r}}]$ belongs to the same category s, while $\mathcal{Z}_d = [\boldsymbol{z_{x_{r+1}}}, \boldsymbol{z_{x_2}}, ..., \boldsymbol{z_{x_n}}]$ are samples from categories other than s. Now, our goal is to shrink the distance between $\boldsymbol{z_{x_i}} \in \mathcal{Z}_s$ and its prototype representation $\boldsymbol{z_{t_i}}$ while enlarging the distance between $\boldsymbol{z_{x_i}}$ and the rest templates.

For this purpose, we optimize on our proposed PCCL loss function (refer to Eq. (6)), where f_{sim} is the sample-wise similarity function. ω_k is a weight parameter ($\omega_k = 1$ for $k \leq r$ and otherwise $\omega_k = \frac{r}{n-r}$), γ is a hyper-parameter for balancing two terms. This first term is a variant of the InfoNCE loss [40]. However, our approach differs from InfoNCE loss by leveraging on the prototypical images that act as anchors for the image categories. In addition, simply optimizing the first term does not necessarily place the prototype image at the class center, thus deviating from our evaluation protocol under CL2N. Therefore, we enforce this property by adding the latent loss, $\sum_{i=1}^{n} \|\boldsymbol{z}_{x_k} - \boldsymbol{z}_{t_k}\|_2$.

$$\begin{aligned} \mathcal{L}_{PCCL} = &-\gamma \log \frac{\sum_{k=1}^{r} \omega_k \exp\left(f_{\text{sim}}\left(\boldsymbol{z}_{x_k}, \boldsymbol{z}_{t_k}\right)\right)}{\sum_{k=1}^{n} \omega_k \exp\left(f_{\text{sim}}\left(\boldsymbol{z}_{x_k}, \boldsymbol{z}_{t_k}\right)\right)} \\ &+ \sum_{k=1}^{n} \|\boldsymbol{z}_{x_k} - \boldsymbol{z}_{t_k}\|_2 \end{aligned} \tag{6}$$

3.4. Deep Supervision by Multi-Task Learning

While it has been shown that a deeper model can contribute to a better generalization, the backbone of VPE is a long chain that is difficult to optimize due to the long path of the gradient flow. We integrate the deep supervision technique with the goal of improving cross-domain generalization and stabilize training. Our idea to create additional paths for gradient flow is inspired by the method from literature [30, 31]. While [30, 31] used a weighted sum of losses from all auxiliary classifiers, we empirically found that better generalization can be achieved when the side-branches are supervised by classification and template reconstruction separately. In other words, we integrated the multi-task learning into the few-shot learning paradigm.

We utilize the DenseNet architecture for class label prediction as its feed-forward connections can alleviate the vanishing gradients problem [20]. This module takes the latent vectors as the input and predicts the class labels. This classification branch is trained in conjunction with the reconstruction task by optimizing a negative log likelihood of the prediction $\mathbf{y}$:

$$\mathcal{L}_{mtl} = -\mathbb{E}_{\mathbf{z_x}\sim p(\mathbf{z_x}|\mathbf{x})}[\log p(\mathbf{y}|\mathbf{z_x})] \tag{7}$$

4. Experiments

In this section, we conducted experiments to validate our approach (VPE++) and compare it against the VPE and other *state-of-the-art* approaches.

Logo Datasets: For the purpose of comparison, we utilized the same five logo datasets and the experimental design followed by Kim et al [27]. These logo datasets include two traffic sign datasets (GTSRB, TT100K) and three logo datasets (Belga, Flickr32, and Toplogos-10). The detailed description of these datasets can be found in literature [27].

Gesture Dataset: We show that gesture prototypes can be used to perform one-shot gesture classification. In this regard, we utilized the American Sign Language dataset [1, 37]

consisting of 26 gesture classes representing 26 English alphabets from A to Z. Each category has 3,000 image samples. The samples are images extracted from a few gesture video clips [37]. The gesture categories corresponding to the alphabets from A to Z, where A to P were used for training and Q to Z were used for testing. The gesture examples and their prototypes were illustrated in the Fig. 5.

Optimization Parameters: The VPE++ model is optimized using the Adam optimizer. We used a batch size of 128 in our experiments. For all experiments except for GTSRB→TT100K, the best performance is achieved by using a learning rate of 1×10^{-4}. We found that the performance is robust to the weight balancing of our loss function and thereby we simply set $\alpha = 1$, $\beta = 1$, and $\gamma = 1$. For the GTSRB→TT100K experiment, the best performance is achieved using a learning rate of 2×10^{-4}, $\alpha = 0.1$, $\beta = 1$, and $\gamma = 1$.

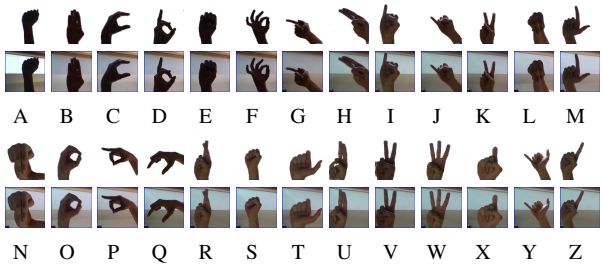


Figure 5. An illustration of gesture prototypes versus real images. The gestures are listed from A to Z.

4.1. Few-Shot Classification on Logo Datasets

First, we compare the results obtained by VPE++ and VPE approaches with respect to *one-shot* recognition accuracies [27]. We performed these experiments using traffic sign datasets and logo datasets. We conducted four experiments to validate our methodology. They are described as follows. 1. Train the network on a sub-set of classes in the GTSRB dataset [48], and test the generalization performance on other GTSRB classes. 2. Train the network on the GTSRB dataset and test the network on the TT100K dataset [62]. 3. Train the network on the Belga logo dataset [24], validate on the Toplogos dataset [49], and test the model performance on the Flickr32 dataset [43]. 4. Train the network on the Belga dataset, validate on the Flickr32 dataset and then test the model on the Toplogos dataset. The data partitions were made following the protocol proposed by Kim et al [27].

Note that a scenario with C unseen categories with K examples for each category is denoted as C-way K-shot classification. Moreover, we conducted experiments in two conditions: 1. *Open set condition*: Test on a combination of seen and unseen categories and 2. *Completely unseen condition*: Test only on unseen categories. While the former condition assumes that the examples present in the testing stage can belong to both seen and new categories, the latter condition assumes that the new examples belong to unseen classes alone. For the purpose of comparison, we reported the *one-shot* accuracies obtained by other algorithms reported by [27] in the Table 1 and 2. The highest accuracies are highlighted as **bold** digits, and the second highest accuracies are colored as blue digits.

Table 1. 1-shot learning accuracy on brand logo datasets.

	Belga–>Flickr32		Belga –>Toplogos	
Split	All	Unseen	All	Unseen
No. classes	32	28	11	6
No. support set	32-way		11-way	
SiamNet	23.25	21.37	37.37	34.92
SiamNet + aug	24.7	22.82	30.84	30.46
QuadNet	40.01	37.72	39.44	36.62
QuadNet + aug	31.68	28.55	38.89	34.16
MatchNet	45.53	40.95	44.35	35.24
MatchNet +aug	38.54	35.28	28.46	27.46
VAE	25.01	25.48	21.9	15.89
VAE+aug	27.7	27.31	23.3	18.59
VPE	28.71	27.34	28.01	26.36
VPE+aug	51.83	50.25	47.48	41.82
VPE+aug+stn	56.6	53.53	58.65	57.75
Ours	**65.54**	**62.56**	**65.57**	**70.27**

Table 2. 1-shot learning accuracy on traffic sign datasets.

	GTSRB –>GTSRB	GTSRB –>TT100K	
Split	Unseen	All	Unseen
No. classes	21	36	32
No. support set	(22+21)-way	36-way	
SiamNet	22.45	22.73	15.28
SiamNet + aug	33.62	28.36	22.74
QuadNet	45.2	42.3	N/A
MatchNet	26.03	53.16	49.53
MatchNet +aug	53.3	62.14	58.75
VAE	20.67	33.14	29.04
VAE+aug	22.24	32.1	27.98
VPE(48*48)	55.3	52.08	49.21
VPE+aug	69.46	66.62	63.91
VPE+aug+stn	74.69	66.88	64.07
VPE(64*64)	56.98	55.58	53.04
VPE+aug	81.27	68.04	64.8
VPE+aug+stn	83.79	73.98	71.8
Ours	**86.51**	**78.31**	**76.36**

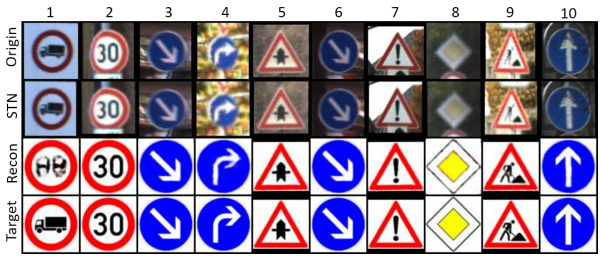
Results show that VPE++ outperforms other approaches in all experimental scenarios by a significant margin. Table 1 shows the one-shot classification accuracies for the brand logos. Consider the Belga->Flickr32 scenario, we achieved 65.54% classification accuracy in the *open set condition* which is 9.94% higher than the second best approach. Furthermore, we achieved a classification accuracy of 62.56% in the *completely unseen condition* which is 9.03% higher than the second best method (VAE+aug+stn). For Belga->Toplogos experiment, we achieved 65.57% *open set condition* classification accuracy and 70.27% in the *completely*

unseen condition, which outperforms the second best method by 6.92% and 12.51% respectively. This is a significant performance boost compared to the existing approaches.

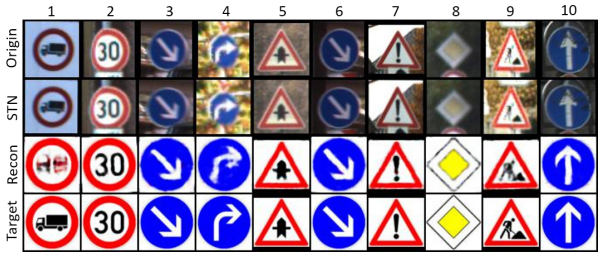
Similarly, our approach also outperforms other methods on the traffic sign dataset, as shown in the Table 2. For *completely unseen condition*, we achieved 86.51% in GTSRB->GTSRB experiment and 76.36% in GTSRB->TT100K experiment, which is 2.71% and 4.56% superior to the second best result, respectively. For *open-set condition*, the VPE++ achieves 78.3%, which is 4.33% better than the second best.

4.1.1 Logo Reconstruction on Unseen Images

Next, we compared the performance of the VPE++ and VPE on a subset of GTSRB road sign images. Fig. 6 depicts the images of true and predicted prototypes. It can be observed that the reconstructed images obtained by VPE++ are visually better in comparison to the ones generated by VPE. This shows that VPE++ model was able to better represent the images in the latent space.



(a) Ours



(b) VPE

Figure 6. A comparison of image reconstruction of: (a) Our approach and (b) VPE model on the unseen road logo images. The images in the 4 rows are: 1) Input image, 2) Input image after STN transformation, 3) Network reconstruction output, and 4) Target prototypical image.

4.2. One-Shot Gesture Recognition on ASL dataset

In addition to validating on logo datasets, the VPE++ approach was tested on the one-shot gesture recognition task using American Sign Language dataset. In this experiment, we have 16 seen classes and 10 unseen classes. First, we computed 1-shot 10-way accuracies as shown in the Table 3. We report the classification accuracy of each class, and the mean class accuracy is obtained by averaging the accuracy of all 10 unseen classes. VPE++ achieves a mean accuracy of 61.28%, outperforming the VPE by 8.31%. On the per-class basis, our approach outperforms VPE for 9 out of 10 gesture classes.

Next, we computed the confusion matrix associated with the ten unseen classes as shown in the Fig. 7. We noticed that most of the errors originate from the samples that are visually similar. For example, gestures related to 'R' and 'U' are similar (fingers crossed) and hence, the classification errors are higher. Similarly, the gestures 'S', 'T', 'X' consist of closed fingers and resemble a fist. These results show that our learned mapping is a meaningful metric for comparing gesture similarity, but the discriminability is low for visually similar gestures.

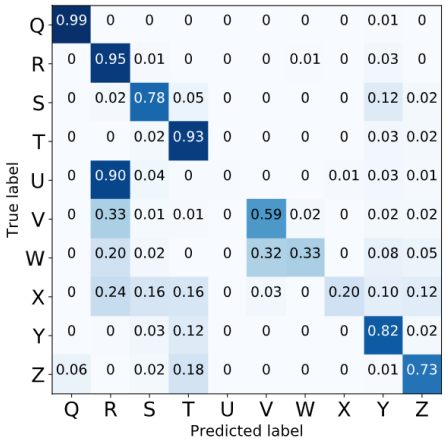


Figure 7. Confusion matrix of one-shot prediction on ASL dataset.

Furthermore, the gesture images are shown in the embedding space, as shown in the Fig. 8. Note that these images are the frames extracted from the gesture video clips. The temporal structure can be observed in the embedding space, where each video clip is shown as a trajectory. It can be noticed that the temporal information is consistent as the gesture sequences appear as continuous trajectories in the latent embedding space.

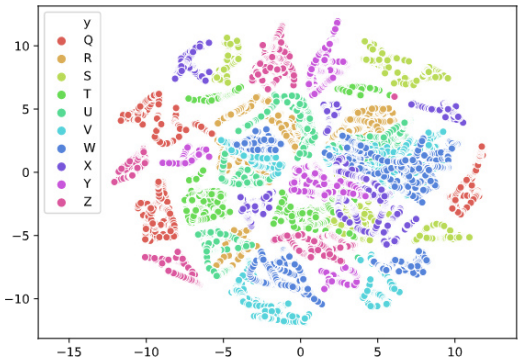


Figure 8. Visualization of the embeddings of the visual sequences, with each gesture class assigned a unique color. We observed that the temporal information is maintained in the latent space.

Table 3. One-shot gesture recognition accuracies on the ASL gesture dataset. Alphabets (Q-Z in this table) indicate the gesture classes.

	mean class accuracy	Q	R	S	T	U	V	W	X	Y	Z
VPE	52.97	95.40	61.13	33.13	74.30	**32.93**	53.06	32.67	7.77	77.2	62.07
Ours	**63.26**	**98.86**	**94.60**	**77.83**	**93.37**	0.00	**58.83**	**33.33**	**19.76**	**82.5**	**73.03**

4.3. Ablation Study

4.3.1 Ablation Study on Model Performance

Next, we conducted an ablation study to show how each improvement is contributing towards an increase in one-shot accuracies for the VPE++ approach on Belga→Flickr32 task. The results obtained from the ablation experiment are shown in the Fig. 4. In this table, *VPE backbone*, *anti-hubness*, *PCCL* and *MTL* refer to the original VPE approach (Sec. 3.1), hubness reduction scheme (Sec. 3.2), contrastive loss (Sec. 3.3) and multi-task loss (Sec. 3.4) respectively.

It can be observed that the hubness reduction techniques have significantly contributed towards the improvement in performance i.e. accuracies improved by 6.51% and 6.39% for *open set* and *completely unseen* conditions, respectively. These results show that VPE is significantly affected by the data hubness problem. Lastly, it is worth noting that the improvements in unseen classification accuracies after introducing PCCL and MTL supervision are relatively small. We conclude that the boost in the performance of VPE++ is achieved due to hubness reduction and the combination of all three loss components.

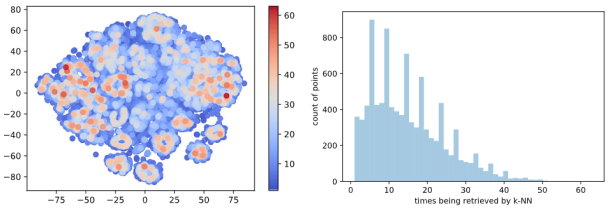
Table 4. An ablation study: step by step improvements of VPE++ approach on Belga →Flickr32

method	①	②	③	④
VPE backbone	✓	✓	✓	✓
Anti-Hubness	✓	✓	✓	
PCCL	✓	✓		
MTL supervision	✓			
All classes	**65.54**	64.71	63.11	56.60
Unseen classes	**62.56**	61.67	59.72	53.53

4.3.2 Anti-Hubness by CL2N

We discovered that CL2N can be used as an effective anti-hubness technique for one-shot learning. We conducted another ablation study to compute the skewness metric [4, 11] on the embedding space created from Flickr32 in Belga→Flickr32 experiment. The metric S^k is given in Eq. (8). The k-occurrence of sample $\boldsymbol{x}$ (denoted as $O^k(\boldsymbol{x})$) means the number of times that sample $\boldsymbol{x}$ is retrieved by k-NN search. μ_{O^k} and σ_{O^k} denote the mean and standard deviation of the k-occurrence distribution. A higher skewness value indicates that the hubness problem is severe and vice versa. The skewness values of raw embedding, normalized raw embedding and mean centered - normalized raw embedding are 1.867, 1.3478 and 1.311, respectively.

$$S^k = \frac{1}{\sigma_{O^k}^3}\mathbb{E}\left[\left(O^k - \mu_{O^k}\right)^3\right] \tag{8}$$



(a) (b)

Figure 9. (a) Visualization of the embedding space after centroid subtraction and normalization. (b) Histogram of samples being retrieved. The long-tail effect was reduced compared to the Fig. 3.

The improvement of centroid subtraction and normalization can also be visualized. We use the same experiment setting as in the Fig. 3 in Sec. 3.2. Intuitively, Fig. 9 (a) shows that the hub points distribute more evenly compared to the Fig. 3 (a). Furthermore, Fig. 9 (b) shows that the long tail effect is partially reduced i.e. there are relatively more samples that are retrieved frequently. In other words, the retrieval rate of the hub points is significantly reduced.

5. Conclusions

This paper proposes an improved one-shot learning algorithm (VPE++) which is built on top of the existing VPE approach. The VPE model was successful in utilizing prototypes to achieve superior performances in the one-shot learning tasks. However, there are several limitations to the existing VPE approach that are addressed in this work. First, we discovered and demonstrated through experiments that VPE suffers from the data hubness problem which hampers its performance on the recognition of novel classes. Hence, a hubness reduction scheme using CL2N metric was incorporated into the VPE++ approach. Second, we proposed a prototype centered contrastive loss function to facilitate contrastive learning by directly using the available prototypical images. Third, we improved the gradient backpropagation of the VPE backbone by utilizing multi-task learning. Experiments were conducted on five logo datasets and a gesture recognition dataset, and the results show that VPE++ considerably outperforms VPE on all of the one-shot recognition tasks.

6. Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant NSF NRI #1925194. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] Robbin Battison. Lexical borrowing in american sign language. 1978. 5

[2] Chris Careaga, Brian Hutchinson, Nathan Hodas, and Lawrence Phillips. Metric-based few-shot learning for video action recognition. *arXiv preprint arXiv:1909.09602*, 2019. 2

[3] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *arXiv preprint arXiv:2002.05709*, 2020. 3, 4, 5

[4] Ali Cheraghian, Shafin Rahman, Dylan Campbell, and Lars Petersson. Mitigating the hubness problem for zero-shot learning of 3d objects. *arXiv preprint arXiv:1907.06371*, 2019. 4, 8

[5] Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. Word translation without parallel data. *arXiv preprint arXiv:1710.04087*, 2017. 2, 3, 5

[6] Judy S DeLoache. Becoming symbol-minded. *Trends in cognitive sciences*, 8(2):66–70, 2004. 2

[7] Georgiana Dinu, Angeliki Lazaridou, and Marco Baroni. Improving zero-shot learning by mitigating the hubness problem. *arXiv preprint arXiv:1412.6568*, 2014. 2

[8] Wei Dong, Charikar Moses, and Kai Li. Efficient k-nearest neighbor graph construction for generic similarity measures. In *Proceedings of the 20th international conference on World wide web*, pages 577–586, 2011. 5

[9] Li Fei-Fei, Rob Fergus, and Pietro Perona. A bayesian approach to unsupervised one-shot learning of object categorie. *IEEE*, 2003. 1, 2

[10] Li Fei-Fei, Rob Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence*, 28(4):594–611, 2006. 2

[11] Roman Feldbauer and Arthur Flexer. A comprehensive empirical comparison of hubness reduction in high-dimensional spaces. *Knowledge and Information Systems*, 59(1):137–166, 2019. 8

[12] Roman Feldbauer, Thomas Rattei, and Arthur Flexer. scikit-hubness: Hubness reduction and approximate neighbor search. *arXiv preprint arXiv:1912.00706*, 2019. 5

[13] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1126–1135. JMLR. org, 2017. 2

[14] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org. 3

[15] Edouard Grave, Armand Joulin, and Quentin Berthet. Unsupervised alignment of embeddings with wasserstein procrustes. *arXiv preprint arXiv:1805.11222*, 2018. 2

[16] Michelle Guo, Edward Chou, De-An Huang, Shuran Song, Serena Yeung, and Li Fei-Fei. Neural graph matching networks for fewshot 3d action recognition. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 653–669, 2018. 2

[17] Yandong Guo and Lei Zhang. One-shot face recognition by promoting underrepresented classes. *arXiv preprint arXiv:1707.05574*, 2017. 2

[18] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. *arXiv preprint arXiv:1911.05722*, 2019. 3

[19] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. *Iclr*, 2(5):6, 2017. 4

[20] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 5

[21] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. In *Advances in neural information processing systems*, pages 2017–2025, 2015. 3

[22] Saumya Jetley, Bernardino Romera-Paredes, Sadeep Jayasumana, and Philip Torr. Prototypical priors: From improving classification to zero-shot learning. *arXiv preprint arXiv:1512.01192*, 2015. 1

[23] Gunnar Johansson and Fredrik Backlund. Drivers and road signs. *Ergonomics*, 13(6):749–759, 1970. 2

[24] Alexis Joly and Olivier Buisson. Logo retrieval with a contrario visual query expansion. In *Proceedings of the 17th ACM international conference on Multimedia*, pages 581–584, 2009. 6

[25] Armand Joulin, Piotr Bojanowski, Tomas Mikolov, Hervé Jégou, and Edouard Grave. Loss in translation: Learning bilingual word mapping with a retrieval criterion. *arXiv preprint arXiv:1804.07745*, 2018. 2

[26] Junsik Kim, Seokju Lee, Tae-Hyun Oh, and In So Kweon. Co-domain embedding using deep quadruplet networks for unseen traffic sign recognition. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018. 2

[27] Junsik Kim, Tae-Hyun Oh, Seokju Lee, Fei Pan, and In So Kweon. Variational prototyping-encoder: One-shot learning with prototypical images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1, 2, 3, 5, 6

[28] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 4

[29] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2, 2015. 2

[30] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu. Deeply-supervised nets. In *Artificial intelligence and statistics*, pages 562–570, 2015. 3, 5

[31] Duo Li and Qifeng Chen. Dynamic hierarchical mimicking towards consistent optimization objectives. *arXiv preprint arXiv:2003.10739*, 2020. 3, 5

[32] F.-F. Li. Knowledge transfer in learning to recognize visual object classes. *International Conference on Development and Learning (ICDL)*, 2006. 2

[33] Junnan Li, Pan Zhou, Caiming Xiong, Richard Socher, and Steven CH Hoi. Prototypical contrastive learning of unsupervised representations. *arXiv preprint arXiv:2005.04966*, 2020. 2, 3, 5

[34] Wen Li, Ying Zhang, Yifang Sun, Wei Wang, Mingjie Li, Wenjie Zhang, and Xuemin Lin. Approximate nearest neighbor search on high dimensional data-experiments, analyses, and improvement. *IEEE Transactions on Knowledge and Data Engineering*, 2019. 5

[35] Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. Meta-sgd: Learning to learn quickly for few-shot learning. *arXiv preprint arXiv:1707.09835*, 2017. 2

[36] Ashish Mishra, Vinay Kumar Verma, M Shiva Krishna Reddy, S Arulkumar, Piyush Rai, and Anurag Mittal. A generative approach to zero-shot and few-shot action recognition. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 372–380. IEEE, 2018. 2

[37] Akash Nagaraj. Image data set for alphabets in the american sign language. https://www.kaggle.com/grassknoted/asl-alphabet, 2018. 5, 6

[38] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018. 2

[39] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018. 3

[40] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018. 4, 5

[41] S. J. Pan and Q. Yang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, Oct. 2010. 2

[42] Miloš Radovanović, Alexandros Nanopoulos, and Mirjana Ivanović. Hubs in space: Popular nearest neighbors in high-dimensional data. *Journal of Machine Learning Research*, 11(Sep):2487–2531, 2010. 2

[43] Stefan Romberg, Lluis Garcia Pueyo, Rainer Lienhart, and Roelof Van Zwol. Scalable logo recognition in real-world images. In *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, pages 1–8, 2011. 6

[44] Edgar Schonfeld, Sayna Ebrahimi, Samarth Sinha, Trevor Darrell, and Zeynep Akata. Generalized zero-and few-shot learning via aligned variational autoencoders. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8247–8255, 2019. 4

[45] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015. 2, 5

[46] Samuel L Smith, David HP Turban, Steven Hamblin, and Nils Y Hammerla. Offline bilingual word vectors, orthogonal transformations and the inverted softmax. *arXiv preprint arXiv:1702.03859*, 2017. 3

[47] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, pages 4077–4087, 2017. 2

[48] Johannes Stallkamp, Marc Schlipsing, Jan Salmen, and Christian Igel. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural networks*, 32:323–332, 2012. 6

[49] Hang Su, Xiatian Zhu, and Shaogang Gong. Deep learning logo detection with data expansion by synthesising context. In *2017 IEEE winter conference on applications of computer vision (WACV)*, pages 530–539. IEEE, 2017. 6

[50] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1199–1208, 2018. 2

[51] Sarah F Taub. *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge University Press, 2001. 2

[52] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. *arXiv preprint arXiv:1906.05849*, 2019. 3

[53] Nenad Tomašev and Dunja Mladenić. Hubness-aware shared neighbor distances for high-dimensional k-nearest neighbor classification. *Knowledge and information systems*, 39(1):89–122, 2014. 3

[54] Nenad Tomašev, Miloš Radovanovic, Dunja Mladenic, and Mirjana Ivanovic. A probabilistic approach to nearest-neighbor classification: Naive hubness bayesian knn. In *Proc. 20th ACM Int. Conf. on Information and Knowledge Management (CIKM)*, pages 2173–2176, 2011. 3

[55] Nenad Tomašev, Miloš Radovanovic, Dunja Mladenic, and Mirjana Ivanovic. A probabilistic approach to nearest-neighbor classification: Naive hubness bayesian knn. In *Proc. 20th ACM Int. Conf. on Information and Knowledge Management (CIKM)*, pages 2173–2176, 2011. 3

[56] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In *Advances in neural information processing systems*, pages 3630–3638, 2016. 2

[57] Yan Wang, Wei-Lun Chao, Kilian Q. Weinberger, and Laurens van der Maaten. Simpleshot: Revisiting nearest-neighbor classification for few-shot learning, 2019. 3, 4

[58] YAQING Wang, J Kwok, LM Ni, and Q Yao. Generalizing from a few examples: A survey on few-shot learning. *arXiv preprint arXiv:1904.05046*, 2019. 1

[59] Yong Wang, Xiao-Ming Wu, Qimai Li, Jiatao Gu, Wangmeng Xiang, Lei Zhang, and Victor OK Li. Large margin few-shot learning. *arXiv preprint arXiv:1807.02872*, 2018. 3

[60] Jian Zhang, Chenglong Zhao, Bingbing Ni, Minghao Xu, and Xiaokang Yang. Variational few-shot learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1685–1694, 2019. 4

[61] Li Zhang, Tao Xiang, and Shaogang Gong. Learning a deep embedding model for zero-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2021–2030, 2017. 3, 4

[62] Zhe Zhu, Dun Liang, Songhai Zhang, Xiaolei Huang, Baoli Li, and Shimin Hu. Traffic-sign detection and classification in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2110–2118, 2016. 6