Comparison Knowledge Translation for Generalizable Image Classification

Zunlei Feng\textsuperscript{1,4,5}, Tian Qiu\textsuperscript{1}, Sai Wu\textsuperscript{1}, Xiaotuan Jin\textsuperscript{3}, Zengliang He\textsuperscript{3}, Mingli Song\textsuperscript{1,4,5}, Huiqiong Wang\textsuperscript{2,*}

\textsuperscript{1}Zhejiang University
\textsuperscript{2}Ningbo Research Institute, Zhejiang University
\textsuperscript{3}Hangzhou Honghua Digital Technology Co., Ltd.
\textsuperscript{4}Shanghai Institute for Advanced Study of Zhejiang University
\textsuperscript{5}Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies
\{zunleifeng,huiqiong_wang\}@zju.edu.cn

Abstract

Deep learning has recently achieved remarkable performance in image classification tasks, which depends heavily on massive annotation. However, the classification mechanism of existing deep learning models seems to contrast to humans’ recognition mechanism. With only a glance at an image of the object even unknown type, humans can quickly and precisely find other same category objects from massive images, which benefits from daily recognition of various objects. In this paper, we attempt to build a generalizable framework that emulates the humans’ recognition mechanism in the image classification task, hoping to improve the classification performance on unseen categories with the support of annotations of other categories. Specifically, we investigate a new task termed Comparison Knowledge Translation (CKT). Given a set of fully labeled categories, CKT aims to translate the comparison knowledge learned from the labeled categories to a set of novel categories. To this end, we put forward a Comparison Classification Translation Network (CCT-Net), which comprises a comparison classifier and a matching discriminator. The comparison classifier is devised to classify whether two images belong to the same category or not, while the matching discriminator works together in an adversarial manner to ensure whether classified results match the truth. Exhaustive experiments show that CCT-Net achieves surprising generalization ability on unseen categories and SOTA performance on target categories.

1 Introduction

In the past decade, deep learning has achieved remarkable performance in the image classification task. However, it usually costs a vast number of annotations to train a practical model in a real scenario. On the contrary, with only a glance at an image of the object even unknown type, humans can...
to the same category or not, while the matching discriminator works together in an adversarial manner to ensure whether classified results match the truth. The comparison classifier only focuses on target categories; meanwhile, the adversarial optimization between the comparison classifier and the matching discriminator will translate the comparison knowledge of source categories into the comparison classifier.

Experiments demonstrate that, with only tens of labeled samples, the proposed CCT-Net achieves close performance on par with fully supervised methods. When trained with fully annotated samples, CCT-Net achieves state-of-the-art performance. The most surprising is that the proposed CCT-Net shows promising generalization ability on novel categories.

Our contribution is therefore introducing a new CKT-Task, in aim to translate the comparison knowledge learned from fully annotated source categories into novel ones with few labels, which emulates the humans’ recognition mechanism in the image classification task. Furthermore, we propose a dedicated solution CCT-Net that comprises a comparison classifier and a matching discriminator. The proposed CCT-Net is evaluated on a broad domain of image datasets, which shows that CCT-Net achieves SOTA classification performance and surprise generalization on novel categories.

2 Related Work

To improve the performance (precision, convergence time, and robustness) of deep classifiers with as few annotations as possible, various classification tasks, including zero/few-shot learning [Lu et al., 2020], transfer learning [Zhuang et al., 2020], distillation learning [Gou et al., 2020], unsupervised learning [Chen et al., 2020], have attracted interest from many researchers. In what follows, we review here two lines of work that are closely related to ours: zero/few-shot learning and transfer learning.

Zero/few-shot learning can be classified into three categories: mode-based, metric-based, and optimization-based methods. Metric-based methods, including SiameseNet [Koch et al., 2015], Match Net [Vinyals et al., 2016], Relation Net [Sung et al., 2018], and Prototype Net [Jake et al., 2017] are most related to ours. SiameseNet [Koch et al., 2015] is composed of weights shared twin CNNs, which accept a pair of samples as inputs, and their outputs at the top layer are combined in order to output a single pairwise similarity score. Prototypical Net [Jake et al., 2017] classified samples of new categories by comparing the Euclidean distance between the representation of the input sample with a learnable class prototype. Match Net [Vinyals et al., 2016] adopted cosine distance to measure similarity between two representations that are encoded with two different encoders. Unlike Prototypical Net and Matching Net, which use the non-parametric Euclidean distance or cosine distance to measure the similarity between pairwise features, Relation Net [Sung et al., 2018] adopted a learnable CNN to measure pairwise similarity, which takes the concatenation of feature maps of two samples as input and outputs their relation score. [Lu et al., 2020] summarized more variants about those networks.

Unlike the above methods, we adopt the adversarial manner to distinguish whether the image pair matches with the predicted similarity score. The major difference is that the comparison knowledge of source categories is translated into the target categories rather than adopting the annotations to supervise the classifier’s training. What’s more, the cross-attention mechanism throughout the whole classification process is adopted to enhance the discriminant ability of the comparison classifier.

Transfer learning can be categorized into three types: inductive, transductive, and unsupervised transfer learning. The idea of inductive transfer learning, including multi-task learning algorithm [Simoes et al., 2018] and self-taught learning [Niyaz et al., 2018], is to increase approximation of the target probability distribution in the target domain given target tasks are different from the source tasks. In the transductive transfer learning technique [Rajesh and Manikathan, 2017], the source domain has a lot of labeled data, while the target domain has no labeled data. Both source tasks and target tasks of transductive transfer learning are similar, whereas there is a difference in the domain only. Unsupervised transfer learning [Siddhant et al., 2018] is the same as inductive, but the main difference is that there is no labeled data in both the source and the target domains.

The differences between transfer learning and CCT-Net contain two aspects: the knowledge type and the way of knowledge transfer. Transfer learning transfers the discriminant ability from source categories into a classification network for target categories, while CCT-Net translates comparison classification ability for a pair image, which is a general classification ability for both source and target categories. For the knowledge transfer way, transfer learning adopts the fine-tune and co-training strategies to transfer the classification ability. The adversarial training strategy in CCT-Net is adopted to translate the comparison classification ability, which brings the advantage that redundant discriminant ability for source categories will not distract the discriminant ability for target categories. Furthermore, CCT-Net has superior generalization ability on novel categories, as shown in Section 5.3 and Table 2.

3 Comparison Knowledge Translation Task

Inspired by the general boundary knowledge [Feng et al., 2021; Cheng et al., 2021] devised for the segmentation task, we introduce the comparison knowledge, the recognition ability to distinguish whether two images belong to the same category. Then, the Comparison Knowledge Translation Task (CKT-Task) is defined as follows. CKT-Task aims to learn a generalizable image classification framework that can quickly and precisely classify novel categories like humans. In CKT-Task, there are assumed to be labeled source dataset $S^m$ that contains $m$ object categories and target dataset $S^n$ of $n$ object categories. The $m$ categories and $n$ categories are disjoint. CKT-Task is supposed to translate the comparison knowledge of $S^m$ into the comparison classifier $F_0$, which is devised for only concentrating on classifying of $n$ categories. The comparison classifier $F_0$ is expected to learn the generalizable distinguishing ability.
4 Method

Deep learning methods usually require sufficient annotations to train a well-behaved classifier. There are vast public classification datasets with plenty of annotations, which have been exploited by transfer learning and distillation learning for improving classification performance on target domains or categories. However, the difference specificity of domains and categories severely limits classification performance’s upper bound. In CKT-Task, we study the comparison knowledge that is generalizable for different categories. To this end, we propose the Comparison Classification Translation Network (CCT-Net) to improve the classification performance on target categories, which draws on the generalizable knowledge from vast public datasets. Fig. 1 depicts the whole framework of CCT-Net comprising a comparison classifier and a matching discriminator. The adversarial training strategy translates the comparison knowledge of the matching discriminator learned from vast source categories into the comparison classifier focusing on targeted categories.

4.1 Comparison Classifier

In CCT-Net, the comparison classifier $F_\theta$ is designed to be a two-branch architecture. Each branch is composed of multiple encoders. The comparison classifier only focuses on the dataset $S^n$ of $n$ target categories. The input of the comparison classifier is a pair of images, which have two cases similar and non-similar, as shown in the left part in Fig. 1.

With a target image pair $(I_1^t, I_2^t) \in S^n$ as input, two branches learn a pair of representations $(r_1^t, r_2^t)$. Then, a fully connected layer predicts the similarity score $s = p(r_1^t - r_2^t)$ with the representation difference $r_1^t - r_2^t$ as input.

The two branches of the comparison classifier $F_\theta$ share the same architecture but have different parameters. Inspired by the humans’ recognition mechanism that distinguishes the image pair from global to local, cross attention is introduced to compare features of two images at each layer. The comparison classifier will compare basic, middle-level, and high-level semantic features as layers go deeper.

4.2 Matching Discriminator

As shown in Fig. 1, the matching discriminator has the same architecture as the comparison classifier. Unlike the comparison classifier, the input of the matching discriminator contains two parts: a pair of images $(I_1^s, I_2^s)$ and a similarity score $p \in \{0, 1\}$. The matching discriminator is expected to distinguish whether the input image pair $(I_1^t, I_2^t)$ matches with the similarity score $p$.

The input image pairs of the matching discriminator include both the source dataset $S^m$ and the target dataset $S^n$. Given a target image pair $(I_1^t, I_2^t) \in S^n$ and the similarity score $p(r_1^t - r_2^t)$ predicted by the comparison classifier $F_\theta$, the matching discriminator first learns a pair of representations $(r_1^t, r_2^t)$. Then, the representation difference $r_1^t - r_2^t$ concatenated with a similarity vector $[p(r_1^t - r_2^t), p(r_1^t - r_2^t), \ldots, p(r_1^t - r_2^t)]$ is input into the fully connected layer of the matching discriminator, which will discriminate whether the similarity of the image pair $(I_1^t, I_2^t)$ matches with the predicted similarity score $p(r_1^t - r_2^t)$.

For the image pair $(I_1^s, I_2^s)$ from the source dataset $S^m$, it is assigned with a similarity score $p^s$ (assigning 0 or 1 in this paper) and annotated with the matching condition $c^s$ (True or False, set as 1 or 0 in the code). Four kinds of input cases for the source category are summarized in Fig. 1. Given the source image pair $(I_1^s, I_2^s) \in S^n$ and the assigned similarity score $p^s$, the matching discriminator first learn a pair of representations $(r_1^s, r_2^s)$. Then, the representation difference $r_1^s - r_2^s$ concatenated with the similarity vector $[p^s, p^s, \ldots, p^s]$ is input into the fully connected layer of the matching discriminator. The annotated matching condition $c^s$ for the triplet $(I_1^s, I_2^s; p^s)$ will supervise the matching discriminator $D_\phi$ to learn the matching discrimination ability with the following
matching loss function:
\[
L_c = -c^s \log(\hat{c}) + (1 - c^s) \log(1 - \hat{c}),
\]
where \( \hat{c} \) denotes the output of the matching discriminator \( D_\phi \).

4.3 Comparison Knowledge Translation

With the annotations of the source dataset, the matching discriminator can learn the matching discrimination ability. Then, the adversarial training strategy [Goodfellow et al., 2014] is adopted to translate the comparison knowledge learned by the matching discriminator \( D_\phi \) into the comparison classifier \( F_\theta \) with the following minimax objective:

\[
\min_{F_\theta} \max_{D_\phi} \mathbb{E}_{(I_1^s, I_2^s, p^s)} \left\{ \log[D_\phi(I_1^s, I_2^s, p^s)] \right\} + \mathbb{E}_{(I_1^t, I_2^t) \sim P_t} \left\{ \log[1 - D_\phi(I_1^t, I_2^t, F_\theta(I_1^t, I_2^t))] \right\},
\]

where \( P_s \) and \( P_t \) denote the pair data distribution of source and target datasets, respectively.

With the above adversarial training strategy, CCT-Net can be trained in an ‘unsupervised’ learning manner (the target dataset doesn’t have any annotation, the source dataset has sufficient annotations). However, if there are annotations in the target dataset, the binary Cross-Entropy loss will accelerate the training process and improve the final classification performance of the comparison classifier.

Overall, there are three loss functions: the adversarial loss Eqn.(2) (for the whole CCT-Net), the matching loss function Eqn.(1) (for the matching discriminator), and the binary Cross-Entropy loss function (optional, for comparison classifier). The complete training algorithm for CCT-Net is summarized in Algorithm 1&2 of the supplements.

5 Experiments

Dataset. In the experiments, we adopt five datasets, including MNIST [LeCun et al., 1998], CIFAR-10 [Krizhevsky, 2009], STL-10 [Adam et al., 2011], Oxford-IIIT Pets [Parkhi et al., 2012], and mini-ImageNet [Jia et al., 2009], to verify the effectiveness of the proposed CKT-Task and CCT-Net. The proposed task needs the disjoint source and target categories. So, categories of each dataset are evenly split into the source and target categories in the comparison experiments.

Network architecture and parameter setting. For each encoder branch of the comparison classifier and the matching discriminator, ViT-B/16 [Dosovitskiy et al., 2021] is adopted as the backbone. Each encoder of CCT-Net comprises 12 attention heads, where 2 attention heads are used for cross-attention between two sub-branches in each layer. The fully connected layer of the comparison classifier and the matching discriminator share the same architecture. The length of the similarity vector \([p^{f_1}, p^{f_2}, \ldots, p^{f_{10}}]\) is 768. More details are given in the supplements.

5.1 Comparing with SOTA Methods

In this section, the proposed method is compared with unsupervised methods: SCAN [Gansbeke et al., 2020] and SimCLR [Chen et al., 2020], few-shot methods: Prototype Net [Jake et al., 2017] and Simple CNAPS [Bateni et al., 2020], semi-supervised methods: MixMatch [David et al., 2019] and FixMatch [Sohn et al., 2020], transfer learning methods: Transfer(10) and Transfer(100), and fully supervised methods: SiameseNet [Koch et al., 2015], VGG-16 [Simonyan et al., 2014], ResNet-50 [He et al., 2016], MobileNetV2 [Sandler et al., 2018], DenseNet-121 [Huang et al., 2017], and ViT-B/16 [Dosovitskiy et al., 2021].

The fully and semi-supervised methods are only trained on the target categories in the experiment. Transfer(10) denotes the ResNet-50 trained with 10 annotated samples of the target category. 80% of the target datasets are used as annotated samples for semi-supervised methods (MixMatch and FixMatch). Five annotated samples of target categories are
Table 2: Generalization results on novel categories. ‘dataset1 $\forall$ dataset2’ denotes CCT-Net is only trained on dataset1 and tested on unseen dataset2. CCT$^{adv}$ and CCT$^{cross}$ denote CCT-Net without discriminator and cross attention, respectively.

Table 3: The translation results between different dataset settings. ‘Source$\rightarrow$Target’ denotes translating knowledge of the source dataset into the comparison classifier for the target dataset. $N_s$ and $N_t$ denote category number of source dataset and target dataset, respectively. ‘Single$\rightarrow$Single’ denotes translating knowledge of ten categories, each of which is randomly selected from mini-ImageNet at each time, into the lion category.

5.3 Generalization Ability on Novel Category
To further verify the generalization ability of CKT-Task, the trained models are directly tested on novel categories (the trained models have never seen before). Table 2 shows the generalization results on three datasets. ‘dataset1 $\forall$ dataset2’ denotes CCT-Net is only trained on the dataset1 (all categories of dataset1 are evenly split into the source and target categories) and then tested on unseen dataset2 directly. Due to the specific network architecture, most existing classification methods can’t be directly tested on a novel category. So, we only compare the proposed method with SiameseNet, which has the same two-branch architecture. ‘SiameseNet(all)’ denotes that SiameseNet is trained with fully annotated samples of dataset1 and then tested on dataset2.

From Table 2, we can see that CCT-Net(all) achieves about double increment than SiameseNet(all) on all datasets (CIFAR-10 $\forall$ STL-10: (+47.6, +55.4), STL-10 $\forall$ CIFAR-10: (+16.6, +28.0), mini-ImageNet $\forall$ CIFAR-10: (+28.2, +26.2), mini-ImageNet $\forall$ STL-10: (+44.4, +47.3)), which demonstrates the surprising generalization ability of CKT-Task and CCT-Net on novel categories. Even without annotations of target categories, CCT-Net(0) still achieves better generalization ability than SiameseNet(all). What’s more, we further verify the importance of each component of CCT-Net on generalization ability. With annotated samples of target categories, CCT$^{cross}$ (CCT-Net without cross attention) achieves lower classification accuracy than CCT$^{adv}$ (CCT-Net without discriminator), which demonstrates that cross attention mechanism contributes more to the generalization ability of CCT-Net.

5.4 Convergence Speed on Novel Category
Except for the excellent generalization ability, CKT-Task and CCT-Net still have the advantage of fast convergence speed, shown in Fig. 2. Due to the different number of model parameters, we adopt ‘Time / parameters’ (training time (seconds) per 1M parameter) to compare the convergence speed of different methods on novel categories. From Fig. 2, we can see that the proposed CCT-Net have a faster convergence speed on novel category than existing methods, which indicates that CCT-Net imitates the humans’ quick recognition ability on novel category.
To verify the effectiveness of CCT-Net’s components, we do ablation study on the false pair samples (case 3 and case 4 in Fig. 1), discriminator, cross attention, input position of the condition $c_l/s$, and the different numbers of labeled samples of target categories. For the ablation study, the mini-ImageNet dataset is adopted. For $\text{CCT}_{\text{false}}$, $\text{CCT}_{\text{adv}}$, $\text{CCT}_{\text{cross}}$ and $\text{CCT}_{\text{head}}$, 20 labeled pairs for each target category are used. From Table 5, we can see that CCT-Net(20) achieves higher scores than other ablative methods, demonstrating all components’ effectiveness. $\text{CCT}_{\text{adv}}$, achieves the lowest score than other ablative methods, which indicates that the adversarial translation strategy is the most critical factor for CKT-Task and CCT-Net.

### Table 5: The ablation study results of CCT-Net.

| Ablation | Accuracy | F1-score |
|----------|----------|----------|
| $\text{CCT}_{\text{false}}$ | 74.9 | 67.8 |
| $\text{CCT}_{\text{adv}}$ | 76.0 | 69.5 |
| $\text{CCT}_{\text{cross}}$ | 89.0 | 89.2 |
| $\text{CCT}_{\text{head}}$ | 71.3 | 72.6 |
| CCT-Net (20) | 89.8 | 90.0 |

The ablation study results on annotated pairs of the target category are given in Table 4, where we can find that 50-annotated-pairs is a critical cut-off point, supplying relatively sufficient guidance. Even without any guidance of labeled pairs of target categories, CCT-Net still achieves 81.4% accuracy score on target categories, verifying the effectiveness of CKT-Task and CCT-Net again.

### 5.6 Ablation Study

To verify the effectiveness of CCT-Net’s components, we do ablation study on the false pair samples (case 3 and case 4 in Fig. 1), discriminator, cross attention, input position of the condition $c_l/s$, and the different numbers of labeled samples of target categories. For the ablation study, the mini-ImageNet dataset is adopted. For $\text{CCT}_{\text{false}}$, $\text{CCT}_{\text{adv}}$, $\text{CCT}_{\text{cross}}$ and $\text{CCT}_{\text{head}}$, 20 labeled pairs for each target category are used. From Table 5, we can see that CCT-Net(20) achieves higher scores than other ablative methods, demonstrating all components’ effectiveness. $\text{CCT}_{\text{adv}}$, achieves the lowest score than other ablative methods, which indicates that the adversarial translation strategy is the most critical factor for CKT-Task and CCT-Net.

### Table 4: The ablation study on annotated samples of CCT-Net.

| Annotations | Accuracy | F1-score |
|-------------|----------|----------|
| 0           | 81.4     | 89.2     |
| 5           | 89.0     | 89.5     |
| 10          | 89.5     | 90.0     |
| 20          | 94.9     | 95.3     |
| 50          | 95.3     | 97.8     |
| all         |          |          |

5.5 Accuracy Trend with Incremental Category

Fig. 3 shows the classification accuracy trend of different methods with incremental category on mini-ImageNet. The experiments are validated on 28 randomly selected categories. In the whole training process, each method is firstly trained with 4 categories. Next, the previous model is trained on cumulative categories (adding 4 categories at each time) when the previous model converges.

From Fig. 3(b)(c), we can see that the fully supervised method and the transfer learning method have descending trends of classification accuracy with incremental category. The fundamental reason is that the learned classification knowledge of the fully supervised method and the transfer learning method is highly associated with the category. When the category number increases, the recognition ability of the deep model for each category will decrease. On the contrary, the proposed CCT-Net and SiameseNet achieve an ascending trend of classification accuracy with incremental category, which verifies that the learned comparison knowledge is generalizable for the classification task. What’s more, in the last three increment phases, all categories for CCT-Net achieve ascending trend (Fig. 3(a)), while ‘categories 5~8’ in the last three increment phases, all categories for CCT-Net achieves descending trend (Fig. 3(d)).

### 6 Conclusion

This paper studies a new Comparison Knowledge Translation Task (CKT-Task), which imitates humans’ quick and precise recognition ability on novel categories. The goal of CKT-Task is to translate the comparison knowledge of source categories into the deep classifier for new categories in the least effort and dependable way. Toward realizing CKT-Task, we introduce the Comparison Classification Translation Network (CCT-Net), which comprises a comparison classifier and a matching discriminator. The comparison classifier is devised to classify whether two images belong to the same category or not, while the matching discriminator works together in an adversarial manner to ensure whether classified results match the truth. With the adversarial training between the comparison classifier and the matching discriminator, the comparison knowledge of massive public source categories is successfully translated into the deep classifier for target categories. There is no special requirement for the source and target categories, which means that all public classification datasets can be used as the source datasets in the proposed CCT-Net. Exhaustive experiments show that CCT-Net achieves impressive generalization ability and SOTA performance on par with existing methods. Surprisingly, CCT-Net also achieves impressive results, including fast convergence speed and high accuracy on novel categories, revealing its superior generalization ability. We will focus on exploring more generalization knowledge and framework on other tasks in the future.
Acknowledgments

This work is supported by National Natural Science Foundation of China (No.62002318), Key Research and Development Program of Zhejiang Province (2020C01023), Zhejiang Provincial Science and Technology Project for Public Welfare (LG21F020020), Fundamental Research Funds for the Central Universities (2021FZXX001-23), Starry Night Science Fund of Zhejiang University Shanghai Institute for Advanced Study (Grant No. SN-ZJU-SIAS-001), and Zhejiang Lab (No.2019KD0AD01/014).

References

[Adam et al., 2011] C. Adam, L. Honglak, and Y. Ng Andrew. An analysis of single layer networks in unsupervised feature learning. AISTATS, 2011.

[Bateni et al., 2020] P. Bateni, R. Goyal, V. Masrani, F. Wood, and L. Sigal. Improved few-shot visual classification. CVPR, 2020.

[Chen et al., 2020] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. ICLR, 2020.

[Cheng et al., 2021] L. Cheng, Z. Feng, X. Wang, Y. Liu, J. Lei, and M. Song. Boundary knowledge translation based reference semantic segmentation. In IJCAI, 2021.

[David et al., 2019] Berthelot David, Carlini Nicholas, Goodfellow Ian, Papernot Nicolas, Oliver Avital, and Raffel Colin. Mixmatch: A holistic approach to semi-supervised learning. NeurIPS, 2019.

[Dosovitskiy et al., 2021] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, and N. Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. ICLR, 2021.

[Feng et al., 2021] Z. Feng, L. Cheng, X. Wang, X. Wang, Y. Liu, X. Du, and M. Song. Visual boundary knowledge translation for foreground segmentation. AAAI, 2021.

[Ganskebeke et al., 2020] W. V. Ganskebeke, S. Vandenhende, S. Georgoulis, M. Proesmans, and L. V. Gool. Scan: Learning to classify images without labels. ECCV, 2020.

[Goodfellow et al., 2014] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Bengio Y. Generative adversarial nets. NeurIPS, 2014.

[Gou et al., 2020] J. Gou, B. Yu, SJ. Maybank, and D. Tao. Knowledge distillation: A survey. ICCV, 2020.

[He et al., 2016] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. CVPR, 2016.

[Huang et al., 2017] G. Huang, Z. Liu, Vdm Laurens, and K. Q. Weinberger. Densely connected convolutional networks. CVPR, 2017.

[Jake et al., 2017] Snell Jake, Swersky Kevin, and S. Zemel Richard. Prototypical networks for few-shot learning. NeurIPS, 2017.

[Jia et al., 2009] D. Jia, D. Wei, R. Socher, L. J. Li, L. Kai, and F. F. Li. Imagenet: A large-scale hierarchical image database. CVPR, 2009.

[Jolles et al., 2010] D. D. Jolles, M. J. Grol, Mav Buchem, Sarb Rombouts, and E. A. Crone. Practice effects in the brain: Changes in cerebral activation after working memory practice depend on task demands. Neuroimage, 52(2):658–668, 2010.

[Koch et al., 2015] G. Koch, R. Zemel, and R. Salakhutdinov. Siamese neural networks for one-shot image recognition. ICML, 2015.

[Krizhevsky, 2009] A. Krizhevsky. Learning multiple layers of features from tiny images. Computer Science, 2009.

[LeCun et al., 1998] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998.

[Lu et al., 2020] J. Lu, P. Gong, J. Ye, and C. Zhang. Learning from very few samples: A survey. arXiv, 2020.

[Niyaz et al., 2018] Q. Niyaz, W. Sun, A. Y. Javaid, and M. Alam. A deep learning approach for network intrusion detection system. Frontiers in Pharmacology, 2018.

[Parkhi et al., 2012] O. M. Parkhi, A. Vedaldi, A. Zisserman, and C. V. Jawahar. Cats and dogs. CVPR, 2012.

[Rajesh and Manikanth, 2017] M. Rajesh and S Manikanth. Annoyed realm outlook taxonomy and multi-task learning in qosr modeling: Advances and challenges. Frontiers in Pharmacology, 2018.

[Siddhant et al., 2018] A. Siddhant, A. Goyal, and A. Metallinou. Unsupervised transfer learning for spoken language understanding in intelligent agents. AAAI, 2018.

[Simoes et al., 2018] R. S. Simoes, V. G. Maltarollo, P. R. Oliveira, and K. M. Honorio. Transfer and multi-task learning in deep neural networks. Neuroimage, 2018.

[Simonyan et al., 2014] K. Simonyan, A. Zisserman, and et al. Very deep convolutional networks for large-scale image recognition. Computer Science, 2014.

[Sohn et al., 2020] K. Sohn, D. Berthelot, C. L. Li, Z. Zhang, N. Carlini, E. D. Cubuk, A. Kurakin, H. Zhang, and C. Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. NeurIPS, 2020.

[Sung et al., 2018] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales. Learning to compare: Relation network for few-shot learning. CVPR, 2018.

[Vinyals et al., 2016] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra. Matching networks for one shot learning. NeurIPS, 2016.

[Zhuang et al., 2020] F. Zhuang, Z. Qi, K. Duan, D. Xi, and Q. He. A comprehensive survey on transfer learning. Proceedings of the IEEE, 2020.