Supplementary Material
Not All Features Matter: Enhancing Few-shot CLIP with Adaptive Prior Refinement

Xiangyang Zhu*[1], Renrui Zhang*†‡[2,3], Bowei He[1], Aojun Zhou[2], Dong Wang[3], Bin Zhao[3], Peng Gao‡[3]

* Equal contribution † Project leader ‡ Corresponding author

[1]City University of Hong Kong [2]The Chinese University of Hong Kong
[3]Shanghai Artificial Intelligence Laboratory

{xiangyzhu6-c, boweihe2-c}@my.cityu.edu.hk,
{zhangrenrui, gaopeng}@pjlab.org.cn

1. Related Work

1.1. Feature Selection

The proposed prior refinement module essentially conducts feature selection along channel dimension, which is a widely acknowledged dimensionality reduction process. In this section, we provide a comprehensive review of the feature selection and its connection with our approach.

Feature selection is employed to minimize the impact of dimensionality on datasets by efficiently collecting a subset of features that accurately describe or define the data [37, 1, 51, 44]. The primary objective of feature selection is to construct a small yet comprehensive subset of features that capture the essential aspects of the input data [16, 15, 47]. Feature selection helps models avoid over-fitting and simplify computation for both training and inference [23, 4]. It also boosts models' interpretability by refining task-specific features. In machine learning, the reduction of the dimensionality and consequently feature selection is one of the most common techniques of noise elimination [12, 34].

Traditional selection methods rely on statistical measures to select features [48]. These methods are independent of the learning algorithm and require less computation. Classical statistical criteria, such as variance threshold, Fisher score, Pearson's correlation [8], Linear Discriminant Analysis (LDA) [5], Chi-square [26], and Mutual Information [17, 38], are commonly used to assess the significance of features.

In deep neural networks, channel pruning is an essential technique for memory size and computation efficiency [34, 9, 3, 41]. Pruning removes redundant parameters or neurons that do not significantly contribute to the accuracy of results. This condition may arise when the weight coefficients are close to zero or are replicated. Traditional pruning approaches such as least absolute shrinkage and selection operator (LASSO) [42, 45], Ridge regression [45], and Optimal Brain Damage (OBD) [28] are widely utilized. In addition, recent efforts also incorporate channel pruning into various visual or language encoders [35, 22, 29, 50].

Compared with them, the proposed adaptive prior refinement approach considers the consistency between vision and language representations to reduce redundancy, and adaptively refine task-aware features for different downstream domains. It not only reduces computation and parameters, but also improves few-shot performance.

1.2. Vision-Language Models

Vision-language (VL) pre-training has provided foundation models for various cross-modality tasks [31, 30]. Existing vision-language models (VLMs) trained on web-scale datasets manifest superior transferability for diverse downstream tasks [11, 14, 2, 49, 7, 32, 36, 33, 24, 33]. For example, BLIP trains a multimodal encoder-decoder network for text-image retrieval, visual question answering, and other cross-modal generation tasks [31]. SLIP integrates self-supervision into VL contrastive learning which guarantees an efficient pre-training [36]. Flamingo reinforces VLMs' few-shot capability to cross-modal tasks via only a few input/output examples [2]. And the recently proposed BLIP-2 efficiently leverages the pre-trained VLMs and conducts generation tasks via a cross-modal transformer [30]. These methods significantly improve the generalization ability of pre-trained models on downstream tasks through large-scale contrastive training. The alignment between VL data has become a time-tested principle to supervise VL training.

After training, the off-the-shelf models exhibit remarkable feature extraction capability. The representative among them is CLIP [40].

After being trained on 400M internet-sourced image-text pairs, CLIP exhibits outstanding capability to align vision-language representations. And it has been widely adopted and applied to classification [52, 18, 46], visual grounding [20, 25], image retrieval [6, 27], semantic segmentation [19, 43], and other tasks with only limited adapting. In this work, we propose a new few-shot framework based on CLIP and it can also be extended to other VLMs.

2. Methods

In this section, we give a detailed derivation for the optimization objective S in Equation 5 and Equation 6 of the main text. The average inter-class similarity of the k^{th} channel S_k is formulated as

$$
\begin{aligned}
S &= \frac{1}{C^2}\sum_{i=1}^{C}\sum_{\substack{j=1\\ j\neq i}}^{C}\delta(\mathbf{x}^i\odot\mathbf{B},\mathbf{x}^j\odot\mathbf{B}) \\
&= \frac{1}{C^2}\sum_{i=1}^{C}\sum_{\substack{j=1\\ j\neq i}}^{C}\sum_{d=1}^{D} x_d^i\cdot x_d^j\cdot B_d \\
&= \frac{1}{C^2}\sum_{i=1}^{C}\sum_{\substack{j=1\\ j\neq i}}^{C}\sum_{k=d_1}^{d_Q} x_k^i\cdot x_k^j \\
&= \sum_{k=d_1}^{d_Q}\left(\frac{1}{C^2}\sum_{i=1}^{C}\sum_{\substack{j=1\\ j\neq i}}^{C} x_k^i\cdot x_k^j\right) \\
&= \sum_{k=d_1}^{d_Q} S_k,
\end{aligned}
\tag{1}
$$

where we use the same notation as the main text.

After that, we visualize the inter-class similarity and variance criteria for all 1024 channels of the ResNet-50 [21] backbone in Figure 1. We conduct the statistic on ImageNet [13] with the textual representation. We set the balance factor λ in Equation 8 of the main paper to 0.7. We sort the channels in ascending order according to the blended criterion J_k. From the figure, if the first $Q = 500$ channels are selected, we observe that the inter-class similarities are small and the variances are large. In addition, we also observe partial channels (around $k = 900$) are not activated because both their S_k and V_k are 0. The visualization demonstrates that the proposed criteria effectively identify redundant channels.

More similarity maps are presented in Figure 2. From the examples, we observe that the refined channels mainly contain information about the objective class, while the rest channels include more ambient noise and redundancy.

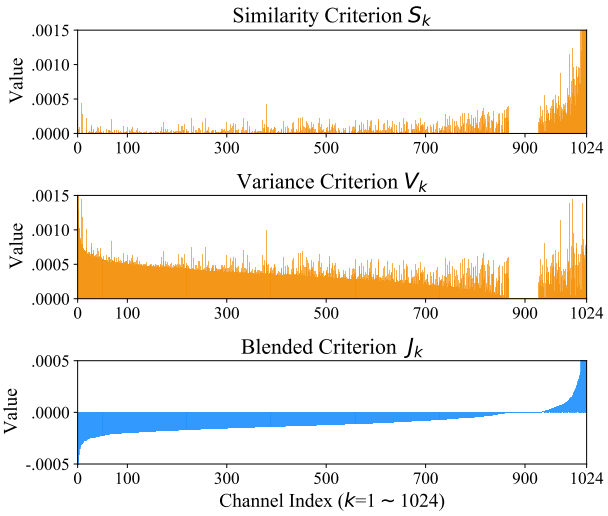


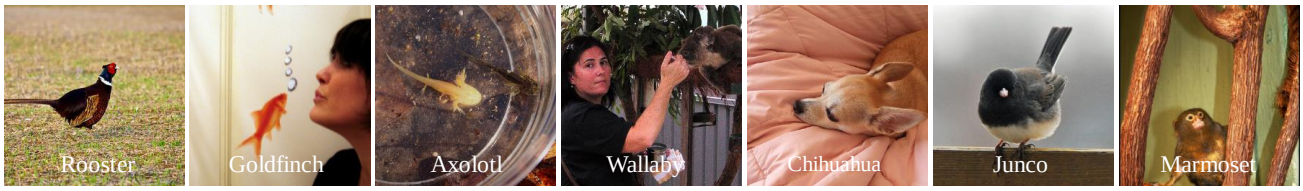
Figure 1: **Visualization of Similarity and Variance Criteria on 1024 Channels** of ResNet-50 encoder.

3. Experiments

Settings. For prior refinement, the number of channels selected, Q, of each dataset is shown in Table 1. For the textual prompt, we follow [46] to ensemble CuPL [39] and template-based prompt [40]. We present the language command for GPT-3 [10] in CuPL prompt generation in Table 2 and 3. The template-based prompt is listed in Table 4.

Channels Number on Vit-B/16. We also verify our channel refinement process on APE with ViT-B/16 backbone as shown in Figure 3 (a), which outputs 512-dimensional representation. This demonstrates the validity of prior refinement in other backbones. For 512-dimensional features, our refinement can also filter out redundancy and noise.

Comparison with PCA. Finally, we compare the proposed significant channel refinement approach with principle component analysis (PCA) in Figure 3 (b), both of which are dimensionality reduction methods. We conduct this experiment on ImageNet with ResNet-50 backbone. We extract $Q = 500$ principal components for the PCA-based approach from the textual representations, similar to our refinement approach. We implement this under both training-free and training-required settings. For training-free variants (denoted as "PCA"), we utilize the transformed representations via PCA to substitute the refined version in APE. For the training-required version (denoted as "PCA-T"), we only optimize the principal components for few-shot learning. The results are presented in Figure 3 (b). We observe that our approach outperforms PCA-based ones in both training-free and training modes. We suggest that this is connected to the change of basis in

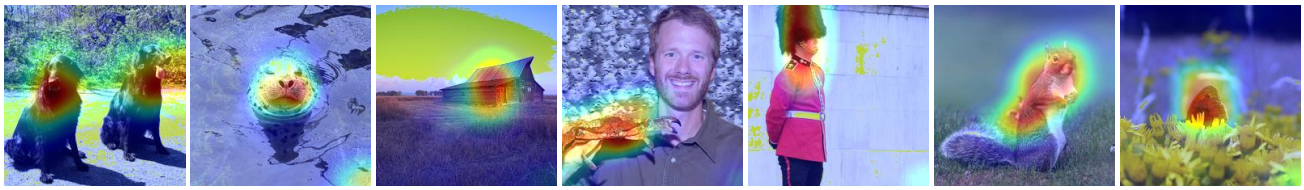


Input Images

(a) Refined Feature Channels

(b) Other Feature Channels

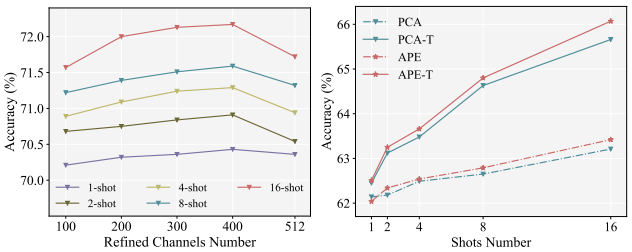
Input Images



(c) Refined Feature Channels

(d) Other Feature Channels

Figure 2: **More Examples of the Similarity Map.** We utilize ResNet-50 [21] visual encoder and refine 512 feature channels from 1024 ones. These examples are collected from ImageNet validation set [13].



(a) Channels Number with ViT-B/16 (b) Comparison with PCA

Figure 3: **More Ablation Study for Prior Refinement.**

Dataset	ImageNet	Caltech-101	DTD	EuroSAT	FGVC
Q	500	900	800	800	900
Food101	Flowers102	Pets	Cars	SUN397	UCF101
800	800	800	500	800	800

Table 1: **Refined Channels Number** Q for each dataset. The backbone is ResNet-50 [21] which extracts 1024-dimensional representations.

PCA algorithm—unexpected biases are introduced to the transformed representations, which is inimical for prediction. As a comparison, our refinement method completely inherits the knowledge from CLIP without transformation, suitable for similarity-based vision-language schemes.

Different Vision-Language Models. We verify our approach on BLIP [31] and SLIP [36], two pre-trained vision-language models. As there are no public results for existing few-shot methods on BLIP and SLIP, we first reproduce their results with official codes on the 11 datasets, and compare the average accuracy for training-free and training-required classification. As shown in Figure 4 and Figure 5, on the two pre-trained models, our APE and APE-T exhibit superior performance and generalization capacity.

Compared to Feature Selection Methods. The proposed Prior Refinement is superior to the sequential feature selection methods. The sequential selection methods highly rely on the selection order while APE does not. We calculate two metrics (similarity and variance) in parallel for each feature channel, and simultaneously select top-Q channels based on the metrics, indicating our method is permutation-invariant to different channels. As we simultaneously and globally select the Q channels, rather than sequentially select, our method is expected to attain an approximate global optimal solution. We also implement forward (FS) and backward (BS) sequential selection algorithms with our metrics in Fig 6. We adopt 10 random seeds and report the average accuracy with standard deviation (shaded area) on ImageNet. As shown, our method performs better than sequential selection methods.

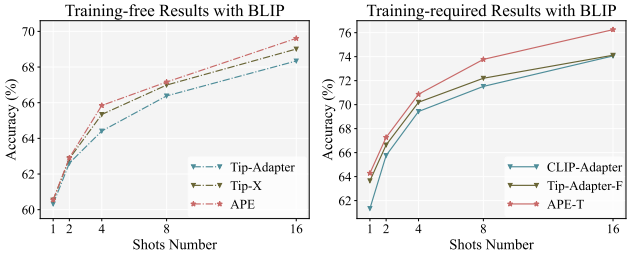


Figure 4: **APE and APE-T on BLIP.**

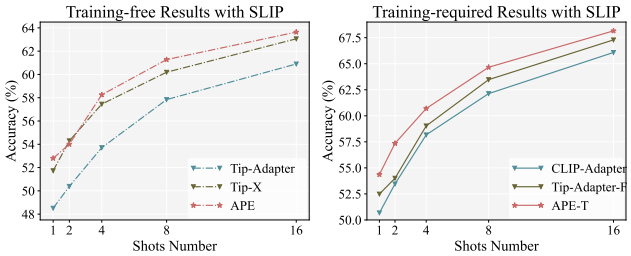


Figure 5: **APE and APE-T on SLIP.**

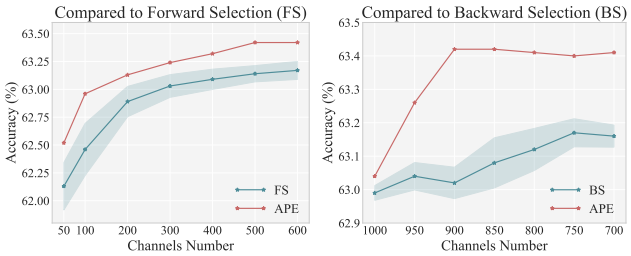


Figure 6: **Comparison with Sequential Feature Selection Methods.**

Dataset	GPT-3 Commands
ImageNet	"Describe what a {} looks like" "How can you identify {}?" "What does {} look like?" "Describe an image from the internet of a {}" "A caption of an image of {}:"
Caltech101	"Describe what a {} looks like" "What does a {} look like" "Describe a photo of a {}"
DTD	"What does a {} material look like?" "What does a {} surface look like?" "What does a {} texture look like?" "What does a {} object look like?" "What does a {} thing look like?" "What does a {} pattern look like?"
EuroSAT	"Describe an aerial satellite view of {}" "How does a satellite photo of a {} look like" "Visually describe a satellite view of a {}"
FGVCAircraft	"Describe a {} aircraft"
Flowers102	"What does a {} flower look like" "Describe the appearance of a {}" "A caption of an image of {}" "Visually describe a {}, a type of flower"
Food101	"Describe what a {} looks like" "Visually describe a {}" "How can you tell the food in the photo is a {}?"
OxfordPets	"Describe what a {} pet looks like" "Visually describe a {}, a type of pet"
StanfordCars	"How can you identify a {}" "Description of a {}, a type of car" "A caption of a photo of a {}:" "What are the primary characteristics of a {}?" "Description of the exterior of a {}" "What are the characteristics of a {}, a car?" "Describe an image from the internet of a {}" "What does a {} look like?" "Describe what a {}, a type of car, looks like"
SUN397	"Describe what a {} looks like" "How can you identify a {}?" "Describe a photo of a {}"
UCF101	"What does a person doing {} look like" "Describe the process of {}" "How does a person {}"

Table 2: **GPT-3 Commands Used in CuPL** (1/2).

Dataset	GPT-3 Commands
ImageNet-V2	"Describe what a {} looks like" "How can you identify {}?" "What does {} look like?" "Describe an image from the internet of a {}" "A caption of an image of {}:"
ImageNet-Sketch	"Describe what a {} looks like" "How can you identify {}?" "What does {} look like?" "Describe an image from the internet of a {}" "A caption of an image of {}:"

Table 3: **GPT-3 Commands Used in CuPL** (2/2).

Dataset	Template Prompt
ImageNet	“itap of a {}.” “a bad photo of the {}.” “a origami {}.” “a photo of the large {}.” “a {} in a video game.” “art of the {}.” “a photo of the small {}.”
Caltech101	“a photo of a {}.”
DTD	“{} texture.”
EuroSAT	“a centered satellite photo of {}.”
FGVCAircraft	“a photo of a {}, a type of aircraft.”
Flowers102	“a photo of a {}, a type of flower.”
Food101	“a photo of {}, a type of food.”
OxfordPets	“a photo of a {}, a type of pet.”
StanfordCars	“a photo of a {}.”
SUN397	“Describe what a {} looks like”
UCF101	“a photo of a person doing {}.”
ImageNet-V2	“itap of a {}.” “a bad photo of the {}.” “a origami {}.” “a photo of the large {}.” “a {} in a video game.” “art of the {}.” “a photo of the small {}.”
ImageNet-Sketch	“itap of a {}.” “a bad photo of the {}.” “a origami {}.” “a photo of the large {}.” “a {} in a video game.” “art of the {}.” “a photo of the small {}.”

Table 4: **Template-based Prompt** for each dataset.

References

[1] Maiwan Bahjat Abdulrazzaq and Jwan Najeeb Saeed. A comparison of three classification algorithms for handwritten digit recognition. In *IEEE International Conference on Advanced Science and Engineering*, pages 58–63, 2019. 1

[2] Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katie Millican, Malcolm Reynolds, et al. Flamingo: a visual language model for few-shot learning. *arXiv preprint arXiv:2204.14198*, 2022. 1

[3] M Augasta and Thangairulappan Kathirvalavakumar. Pruning algorithms of neural networks—a comparative study. *Open Computer Science*, 3(3):105–115, 2013. 1

[4] Shaeela Ayesha, Muhammad Kashif Hanif, and Ramzan Talib. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion*, 59:44–58, 2020. 1

[5] Suresh Balakrishnama and Aravind Ganapathiraju. Linear discriminant analysis-a brief tutorial. *Institute for Signal and information Processing*, 18(1998):1–8, 1998. 1

[6] Alberto Baldrati, Marco Bertini, Tiberio Uricchio, and Alberto Del Bimbo. Effective conditioned and composed image retrieval combining clip-based features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 21466–21474, 2022. 2

[7] Hangbo Bao, Wenhui Wang, Li Dong, Qiang Liu, Owais Khan Mohammed, Kriti Aggarwal, Subhojit Som, and Furu Wei. Vlmo: Unified vision-language pre-training with mixture-of-modality-experts. *arXiv preprint arXiv:2111.02358*, 2021. 1

[8] Jacek Biesiada and Wlodzisław Duch. Feature selection for high-dimensional data—a pearson redundancy based filter. In *Computer recognition systems 2*, pages 242–249, 2007. 1

[9] Davis Blalock, Jose Javier Gonzalez Ortiz, Jonathan Frankle, and John Guttag. What is the state of neural network pruning? *Proceedings of Machine Learning and Systems*, 2:129–146, 2020. 1

[10] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020. 2

[11] Xi Chen, Xiao Wang, Soravit Changpinyo, AJ Piergiovanni, Piotr Padlewski, Daniel Salz, Sebastian Goodman, Adam Grycner, Basil Mustafa, Lucas Beyer, et al. Pali: A jointly-scaled multilingual language-image model. *arXiv preprint arXiv:2209.06794*, 2022. 1

[12] Manoranjan Dash and Huan Liu. Feature selection for classification. *Intelligent data analysis*, 1(1-4):131–156, 1997. 1

[13] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. 2, 3

[14] Yuqing Du, Ksenia Konyushkova, Misha Denil, Akhil Raju, Jessica Landon, Felix Hill, Nando de Freitas, and Serkan Cabi. Vision-language models as success detectors. *arXiv preprint arXiv:2303.07280*, 2023. 1

[15] Adel Sabry Eesa, Adnan Mohsin Abdulazeez Brifcani, and Zeynep Orman. Cuttlefish algorithm-a novel bio-inspired optimization algorithm. *International Journal of Scientific & Engineering Research*, 4(9):1978–1986, 2013. 1

[16] Adel Sabry Eesa, Zeynep Orman, and Adnan Mohsin Abdulazeez Brifcani. A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems. *Expert systems with applications*, 42(5):2670–2679, 2015. 1

[17] Pablo A Estévez, Michel Tesmer, Claudio A Perez, and Jacek M Zurada. Normalized mutual information feature selection. *IEEE Transactions on Neural Networks*, 20(2):189–201, 2009. 1

[18] Peng Gao, Shijie Geng, Renrui Zhang, Teli Ma, Rongyao Fang, Yongfeng Zhang, Hongsheng Li, and Yu Qiao. Clip-adapter: Better vision-language models with feature adapters. *arXiv preprint arXiv:2110.04544*, 2021. 2

[19] Golnaz Ghiasi, Xiuye Gu, Yin Cui, and Tsung-Yi Lin. Open-vocabulary image segmentation. *arXiv preprint arXiv:2112.12143*, 2021. 2

[20] Huy Ha and Shuran Song. Semantic abstraction: Open-world 3D scene understanding from 2D vision-language models. In *Proceedings of the 2022 Conference on Robot Learning*, 2022. 2

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 2, 3, 4

[22] Hengyuan Hu, Rui Peng, Yu-Wing Tai, and Chi-Keung Tang. Network trimming: A data-driven neuron pruning approach towards efficient deep architectures. *arXiv preprint arXiv:1607.03250*, 2016. 1

[23] Xuan Huang, Lei Wu, and Yinsong Ye. A review on dimensionality reduction techniques. *International Journal of Pattern Recognition and Artificial Intelligence*, 33(10):1950017, 2019. 1

[24] Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. Scaling up visual and vision-language representation learning with noisy text supervision. In *International Conference on Machine Learning*, pages 4904–4916, 2021. 1

[25] Haojun Jiang, Yuanze Lin, Dongchen Han, Shiji Song, and Gao Huang. Pseudo-q: Generating pseudo language queries for visual grounding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15513–15523, 2022. 2

[26] Xin Jin, Anbang Xu, Rongfang Bie, and Ping Guo. Machine learning techniques and chi-square feature selection for cancer classification using sage gene expression profiles. In *Data Mining for Biomedical Applications: PAKDD Workshop*, pages 106–115, 2006. 1

[27] Benno Krojer, Vaibhav Adlakha, Vibhav Vineet, Yash Goyal, Edoardo Ponti, and Siva Reddy. Image retrieval from contextual descriptions. *arXiv preprint arXiv:2203.15867*, 2022. 2

[28] Yann LeCun, John Denker, and Sara Solla. Optimal brain damage. *Advances in Neural Information Processing Systems*, 2, 1989. 1

[29] Namhoon Lee, Thalaiyasingam Ajanthan, and Philip HS Torr. Snip: Single-shot network pruning based on connection sensitivity. *arXiv preprint arXiv:1810.02340*, 2018. 1

[30] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. *arXiv preprint arXiv:2301.12597*, 2023. 1

[31] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International Conference on Machine Learning*, pages 12888–12900. PMLR, 2022. 1, 4

[32] Junnan Li, Ramprasaath Selvaraju, Akhilesh Gotmare, Shafiq Joty, Caiming Xiong, and Steven Chu Hong Hoi. Align before fuse: Vision and language representation learning with momentum distillation. *Advances in Neural Information Processing Systems*, 34:9694–9705, 2021. 1

[33] Yangguang Li, Feng Liang, Lichen Zhao, Yufeng Cui, Wanli Ouyang, Jing Shao, Fengwei Yu, and Junjie Yan. Supervision exists everywhere: A data efficient contrastive language-image pre-training paradigm. *arXiv preprint arXiv:2110.05208*, 2021. 1

[34] Tailin Liang, John Glossner, Lei Wang, Shaobo Shi, and Xiaotong Zhang. Pruning and quantization for deep neural network acceleration: A survey. *Neurocomputing*, 461:370–403, 2021. 1

[35] Zhuang Liu, Jianguo Li, Zhiqiang Shen, Gao Huang, Shoumeng Yan, and Changshui Zhang. Learning efficient convolutional networks through network slimming. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2736–2744, 2017. 1

[36] Norman Mu, Alexander Kirillov, David Wagner, and Saining Xie. Slip: Self-supervision meets language-image pre-training. *arXiv preprint arXiv:2112.12750*, 2021. 1, 4

[37] D Lakshmi Padmaja and B Vishnuvardhan. Comparative study of feature subset selection methods for dimensionality reduction on scientific data. In *IEEE International Conference on Advanced Computing*, pages 31–34, 2016. 1

[38] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005. 1

[39] Sarah Pratt, Rosanne Liu, and Ali Farhadi. What does a platypus look like? generating customized prompts for zero-shot image classification. *arXiv preprint arXiv:2209.03320*, 2022. 2

[40] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763, 2021. 2

[41] Russell Reed. Pruning algorithms-a survey. *IEEE Transactions on Neural Networks*, 4(5):740–747, 1993. 1

[42] Fadil Santosa and William W Symes. Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, 7(4):1307–1330, 1986. 1

[43] Gyungin Shin, Weidi Xie, and Samuel Albanie. Reco: Retrieve and co-segment for zero-shot transfer. *arXiv preprint arXiv:2206.07045*, 2022. 2

[44] Jiliang Tang, Salem Alelyani, and Huan Liu. Feature selection for classification: A review. *Data classification: Algorithms and applications*, page 37, 2014. 1

[45] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996. 1

[46] Vishaal Udandarao, Ankush Gupta, and Samuel Albanie. Sus-x: Training-free name-only transfer of vision-language models. *arXiv preprint arXiv:2211.16198*, 2022. 2

[47] S Velliangiri, SJPCS Alagumuthukrishnan, et al. A review of dimensionality reduction techniques for efficient computation. *Procedia Computer Science*, 165:104–111, 2019. 1

[48] B Venkatesh and J Anuradha. A review of feature selection and its methods. *Cybernetics and information technologies*, 19(1):3–26, 2019. 1

[49] Wenhui Wang, Hangbo Bao, Li Dong, Johan Bjorck, Zhiliang Peng, Qiang Liu, Kriti Aggarwal, Owais Khan Mohammed, Saksham Singhal, Subhojit Som, et al. Image as a foreign language: Beit pretraining for all vision and vision-language tasks. *arXiv preprint arXiv:2208.10442*, 2022. 1

[50] Runxin Xu, Fuli Luo, Chengyu Wang, Baobao Chang, Jun Huang, Songfang Huang, and Fei Huang. From dense to sparse: Contrastive pruning for better pre-trained language model compression. *arXiv preprint arXiv:2112.07198*, 2021. 1

[51] Rizgar Zebari, Adnan Abdulazeez, Diyar Zeebaree, Dilovan Zebari, and Jwan Saeed. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *J. Appl. Sci. Technol. Trends*, 1(2):56–70, 2020. 1

[52] Renrui Zhang, Rongyao Fang, Wei Zhang, Peng Gao, Kunchang Li, Jifeng Dai, Yu Qiao, and Hongsheng Li. Tip-adapter: Training-free clip-adapter for better vision-language modeling. *arXiv preprint arXiv:2111.03930*, 2021. 2