PVeRA: Probabilistic Vector-Based Random Matrix Adaptation

Supplementary Material

A. Grid Search

Table 4 shows the values of grid search used for the rank, the reduction ratio, and the learning rate, as well as the percentage of configurations for which each value performed best.

B. Pseudocode for PVeRA

Algorithm 2 shows pseudocode for the initialization and the forward pass through PVeRA. Our Python code is available here.

C. Uncertainty Estimation

Sampling from the learned distribution during inference allows to generate confidence intervals for the predictions by making multiple consecutive passes with each sample. In Figure 9, we look at each dataset of VTAB-1k and analyze images that were correctly classified and incorrectly classified. The predicted class is the most predicted class among the k predictions, and the confidence intervals is computed from the softmax scores of this predicted class. Figure 8 shows the widths of the Monte Carlo confidence intervals for the correctly and incorrectly classified samples across all datasets. It is noteworthy to observe that the confidence intervals of the wrongly classified images is much larger than those of the correctly classified images. Moreover, even when an image is correctly classified, the spread of the confidence interval can indicate uncertainty. For example in Figure 9.a, EuroSAT's image of the `forest` could have easily been classified `river` or `sea & lake`, which explains the higher confidence interval than Caltech101's image of `pizza` for which the uncertainty is much lower.

	32	64	128
LoRA	26%	35%	39%
DoRA	18%	37%	46%

(a) Rank

	16	8	4
Bottleneck	11%	21%	68%
AdaptFormer	11%	23%	67%

(b) Reduction ratio

	10^{-3}	10^{-4}	10^{-5}
Linear	79%	21%	0%
$(IA)^3$	60%	30%	10%
VeRA	46%	23%	32%
PVeRA	67%	28%	5%

(c) Learning rate

Table 4. **Grid search results**. Grid search hyperparameters for (a) ranks, (b) reduction ratios, and (c) learning rates, along with the percentage of the 57 configurations (19 datasets across three seeds) for which each value was chosen.

Algorithm 2. **Initialization and forward pass of PVeRA**. Algorithm for the initialization and forward pass of PVeRA.

Input: $\boldsymbol{x}$, α, $\boldsymbol{A}$, $\boldsymbol{B}$, linear_layer, training, r, d
Initialization:
  $\boldsymbol{b} \leftarrow \text{zeros}(d)$
  $\boldsymbol{d} \leftarrow \text{ones}(2r) \cdot \text{uniform}(10^{-5}, 1)$
Function:
  $\boldsymbol{b} \leftarrow \text{repeat}(\boldsymbol{b}, \boldsymbol{x}.\text{batch\_size}, \boldsymbol{x}.\text{seq\_length})$
  $\boldsymbol{d} \leftarrow \text{repeat}(\boldsymbol{d}, \boldsymbol{x}.\text{batch\_size}, \boldsymbol{x}.\text{seq\_length})$
  $\boldsymbol{\mu}, \boldsymbol{\sigma} \leftarrow \boldsymbol{x} \cdot \boldsymbol{A} \odot \boldsymbol{d}$
  if training **then**
    $\boldsymbol{\epsilon} \leftarrow \text{random\_normal}(\boldsymbol{\mu}.\text{shape})$
    $\boldsymbol{z} \leftarrow \alpha \cdot (\boldsymbol{\mu} + \boldsymbol{\epsilon} \odot \exp(\boldsymbol{\sigma}^2))$
  else
    $\boldsymbol{z} \leftarrow \boldsymbol{\mu}$
  end if
  $\boldsymbol{x}_{\text{adapted}} \leftarrow \alpha \cdot \boldsymbol{z} \odot \boldsymbol{b}$
Returns:
  $\text{linear\_layer}(\boldsymbol{x}) + \boldsymbol{x}_{\text{adapted}}$

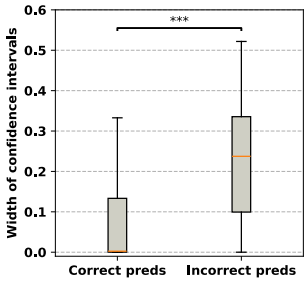
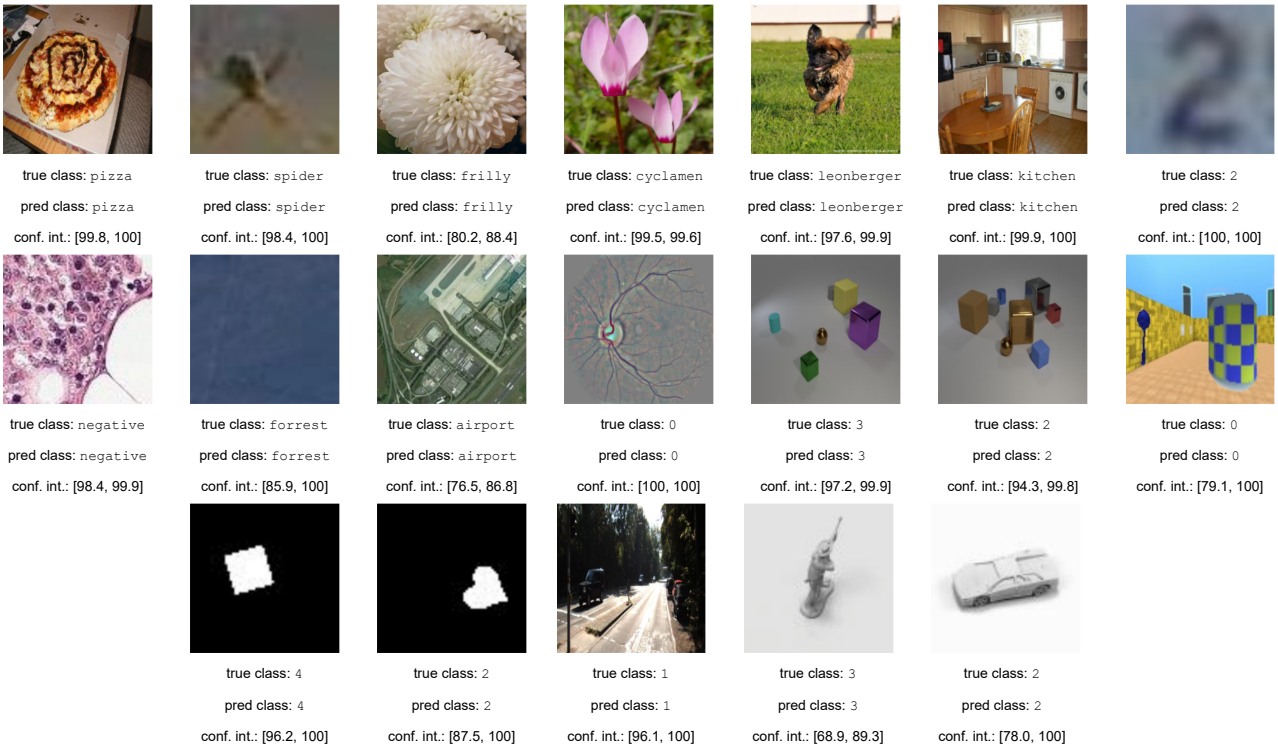
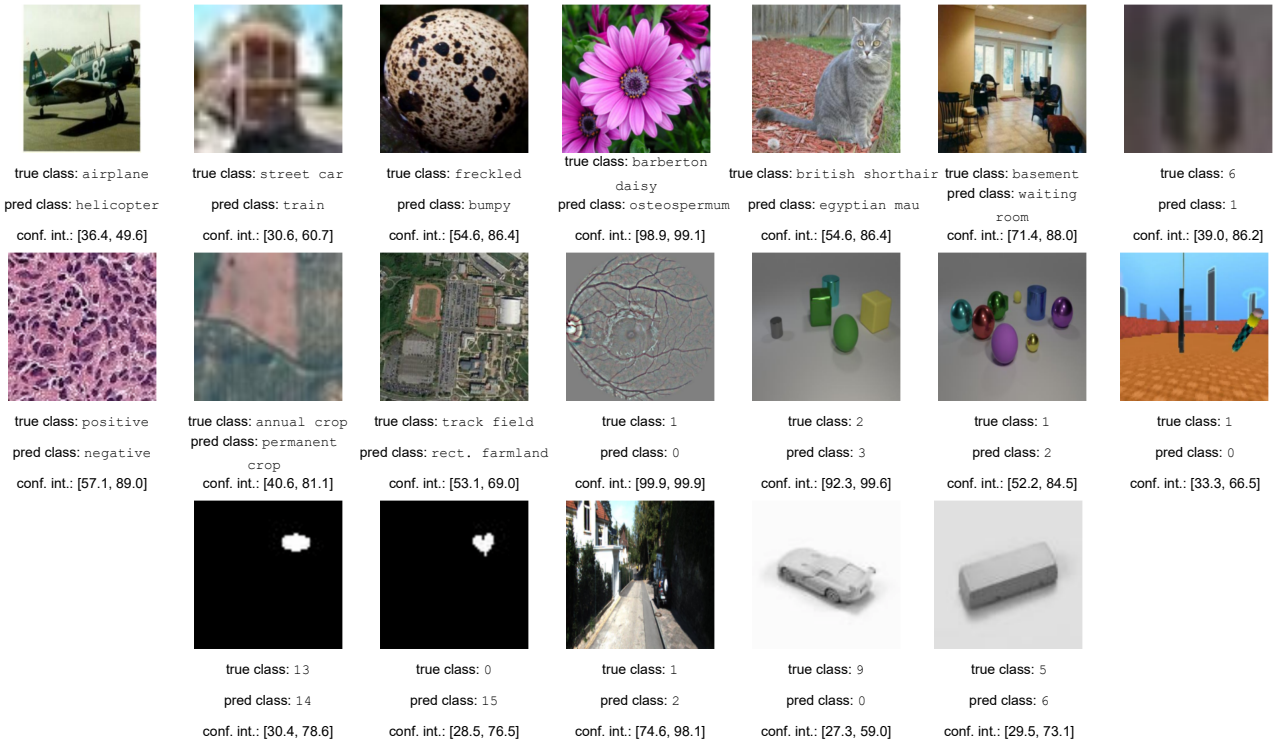


Figure 8. **Width of confidence intervals for correctly and incorrectly classified samples**. Width ((upper bound) − (lower bound)) for correctly classified and incorrectly classified samples across the test set from all VTAB-1k datasets, generated using Monte Carlo inference with random sampling of the PVeRA adapters. The significance level corresponds to the p-value of a one-sided unpaired Wilcoxon test.



(a) Correctly classified



(b) Incorrectly classified

Figure 9. **Confidence interval estimation**. True class, predicted class, and estimated 95% confidence intervals for all 19 datasets (a) correctly classified samples, and (b) incorrectly classified samples.

D. Application on Natural Language

While the main focus of the study is to introduce PVeRA on vision tasks, we include here a small experiment on adapting language models for different tasks. We utilize a DeBERTa-V3-base [15] model, which we adapt to three tasks from the GLUE benchmark [49] (we select the tasks with a low number of samples for reasons of computational efficiency). Table 5 summarizes these results. Even across language classification tasks, PVeRA is very competitive compared to other adapters. While this is not an in-depth study, it shows very promising results. We train across three random seeds, we report results on the initial validation set, and split the provided train set into a training and validation set, as the test set labels are not made available. Note that AdaptFormer is not present as it was introduced for vision tasks and has, to the best of our knowledge, not been used for language tasks.

Adapter	RTE	MRPC	CoLA
Linear	63.2*	68.6*	48.3*
Bottleneck	68.6*	85.5	43.1*
IA^3	<u>76.3</u>	85.6	<u>65.0</u>
DoRA	71.8*	85.5	62.9
LoRA	76.1	85.4	61.3*
VeRA	72.8*	**85.9**	62.4*
PVeRA	**76.7**	<u>85.7</u>	**65.6**

Table 5. **Natural language processing results**. Results on three classification tasks from the GLUE benchmark. All reported metrics are averaged over three random seeds. RTE and MRPC report accuracies, while CoLA reports the Matthews correlation coefficient. Results marked with * indicate accuracies significantly lower ($p = 0.05$) than the accuracy of PVeRA.

E. Detailed Results

Table 6 presents a detailed version of Table 1 with standard deviation of accuracies. We take the binary accuracy of each test set samples, perform a binomial test between each adapter and PVeRA, and mark the results with significantly lower accuracy with *. Table 7 presents similar results for a smaller DINOv2 (ViT-S/14). Conclusions are similar than for a ViT-B/14, with larger gaps in performance, with PVeRA's performance being 67.5% with the second best performance being for LoRA and DoRA at 66.1% and VeRA average performance at 65.9%. For a larger model, DINOv2 (ViT-L/14) presented in Table 8, the gap between the performance of PVeRA (73.3%) and the second best (LoRA with 73.1%) is smaller. We believe that this is due to the fact that with larger models, the performance gets closer to the highest attainable performance, and the gains in performance are more difficult to obtain. Note that for computational efficiency constraints, we did not perform grid search for the ViT-L/14, and instead fixed the hyperparameters to the most chosen hyperparameter reported in Table 4 (e.g., rank of 128 for LoRA, and learning rate of 10^{-3} for VeRA).

	Linear	Bottleneck	IA^3	AdaptFormer	DoRA	LoRA	VeRA	PVeRA
• Caltech101	85.1* ± 0.4	88.6 ± 0.6	86.8* ± 0.6	87.9* ± 1.1	90.0 ± 1.3	90.1 ± 1.7	87.3* ± 1.1	88.7 ± 0.7
• CIFAR-100	61.6* ± 0.2	41.8* ± 11.4	71.3 ± 0.8	56.8* ± 1.9	61.8* ± 0.8	64.5* ± 1.5	70.1* ± 0.0	71.7 ± 0.3
• DTD	73.7* ± 0.3	72.3* ± 1.0	77.2 ± 0.5	72.9* ± 0.9	74.3* ± 0.2	75.3 ± 0.1	76.6 ± 0.3	76.1 ± 1.8
• Flowers102	99.7 ± 0.0	98.7* ± 0.4	99.7 ± 0.0	98.5* ± 0.2	99.1* ± 0.1	99.5* ± 0.0	99.7 ± 0.0	99.7 ± 0.0
• Pets	94.0 ± 0.1	85.8* ± 7.5	94.1 ± 0.1	90.5* ± 0.2	90.7* ± 0.6	91.8* ± 0.4	94.1 ± 0.1	93.3 ± 0.9
• Sun397	51.3* ± 0.7	32.5* ± 21.7	54.2* ± 0.3	30.7* ± 20.4	49.1* ± 0.9	50.7* ± 0.9	54.5* ± 0.2	54.7 ± 0.1
• SVHN	38.9* ± 1.5	90.9 ± 0.8	68.7* ± 1.5	89.3 ± 0.8	91.0 ± 0.7	87.7* ± 0.4	87.8* ± 0.7	88.7 ± 0.3
• Camelyon	82.3* ± 0.6	86.4 ± 0.7	83.2* ± 1.0	86.5 ± 1.4	87.2 ± 0.7	84.7* ± 0.1	84.3* ± 0.5	85.0 ± 0.9
• EuroSAT	90.6* ± 0.2	94.3 ± 0.1	93.0* ± 0.3	94.3 ± 0.5	95.5 ± 0.3	94.9 ± 0.4	93.6* ± 0.4	94.3 ± 0.4
• Resisc45	78.8* ± 0.2	82.7* ± 0.9	84.9* ± 0.5	83.6* ± 1.0	87.7 ± 0.5	84.9* ± 0.7	84.9* ± 0.2	86.4 ± 0.3
• Retinopathy	73.8* ± 0.3	73.6* ± 0.0	74.4* ± 1.1	73.6* ± 0.0	73.6* ± 0.0	75.0 ± 1.1	75.0 ± 0.1	74.8 ± 0.9
• Clevr-Count	42.9* ± 0.8	78.6 ± 1.0	55.9* ± 1.5	86.7 ± 0.7	63.7* ± 2.3	66.2* ± 1.8	58.1* ± 2.5	71.5 ± 4.8
• Clevr-Dist	33.7* ± 1.8	60.5 ± 0.7	51.8* ± 3.7	61.7 ± 1.0	60.5 ± 1.1	59.5* ± 1.1	58.5* ± 1.2	60.6 ± 0.4
• DMLab	41.2* ± 0.8	48.1* ± 3.6	43.3* ± 0.5	49.3 ± 0.8	52.6 ± 1.6	50.3 ± 0.5	47.4* ± 1.2	48.6 ± 0.2
• dSpr-Loc	11.8* ± 3.0	80.4 ± 2.7	55.5* ± 5.5	69.3* ± 4.0	84.3 ± 0.3	76.8 ± 0.5	73.8 ± 3.6	72.1 ± 1.5
• dSpr-Ori	31.0* ± 2.8	49.9 ± 0.8	53.1 ± 0.8	52.2 ± 2.4	52.1 ± 0.2	52.3 ± 0.4	48.0* ± 0.7	49.5 ± 2.1
• KITTI-Dist	54.0* ± 3.6	79.7* ± 2.3	77.7* ± 0.5	82.7 ± 2.0	82.1 ± 1.0	80.8* ± 2.4	84.0 ± 1.7	83.3 ± 0.4
• sNORB-Azim	12.4* ± 0.5	19.1* ± 1.3	14.4* ± 1.7	19.6* ± 1.3	21.2 ± 1.0	19.6* ± 0.8	19.0* ± 0.4	20.3 ± 1.1
• sNORB-Elev	25.5* ± 0.6	31.7* ± 2.5	26.8* ± 0.6	36.7 ± 1.0	32.2* ± 1.2	34.3* ± 2.1	32.3* ± 0.6	37.0 ± 2.8

Table 6. **Detailed VTAB-1k benchmark results using a ViT-B/14**. Average and standard deviation accuracies across three seeds for seven adapters across all 19 VTAB-1k datasets. Results marked with * indicate accuracies significantly lower ($p = 0.05$) than the accuracy of PVeRA.

	Linear	Bottleneck	IA^3	AdaptFormer	DoRA	LoRA	VeRA	PVeRA
• Caltech101	85.9* ± 0.4	62.9* ± 2.1	87.8 ± 1.9	76.2* ± 1.8	86.9* ± 2.1	87.6* ± 1.1	85.4* ± 0.4	88.0 ± 0.2
• CIFAR-100	48.4* ± 0.3	10.2* ± 3.3	59.2 ± 0.4	24.7* ± 2.4	47.9* ± 1.2	46.4* ± 0.4	58.2* ± 0.1	58.5 ± 0.4
• DTD	70.8* ± 0.5	54.3* ± 2.1	72.8 ± 0.7	60.5* ± 1.8	68.1* ± 0.6	68.4* ± 0.2	73.0 ± 0.7	72.6 ± 0.3
• Flowers102	99.3 ± 0.0	76.7* ± 2.0	99.1 ± 0.2	86.5* ± 1.9	95.3* ± 0.2	96.7* ± 0.3	99.2 ± 0.0	99.2 ± 0.1
• Pets	91.6 ± 0.3	37.0* ± 0.9	91.6 ± 0.3	68.0* ± 1.5	84.3* ± 0.8	85.1* ± 1.9	91.9 ± 0.5	91.8 ± 0.4
• Sun397	47.1 ± 0.2	1.1* ± 0.3	47.2 ± 0.0	1.0* ± 0.0	36.3* ± 0.2	38.3* ± 1.4	47.7 ± 0.2	47.4 ± 1.1
• SVHN	35.1* ± 0.7	78.9* ± 0.8	74.3* ± 8.8	82.6* ± 2.5	88.6 ± 0.4	85.8 ± 1.7	82.7* ± 3.1	85.2 ± 0.8
• Camelyon	81.0* ± 0.3	76.9* ± 0.7	85.1 ± 1.1	81.8* ± 1.4	85.7 ± 0.8	85.2 ± 0.5	82.8* ± 1.2	84.9 ± 0.9
• EuroSAT	90.7* ± 0.5	88.2* ± 2.2	94.8 ± 0.5	92.2* ± 1.0	94.8 ± 0.3	93.5 ± 0.6	92.6* ± 0.8	93.0 ± 0.3
• Resisc45	72.9* ± 0.8	66.5* ± 1.3	83.5 ± 0.3	74.4* ± 0.8	82.5 ± 1.0	79.6* ± 0.8	80.0* ± 1.5	82.0 ± 0.6
• Retinopathy	73.6* ± 0.0	73.6* ± 0.0	73.6* ± 0.0	73.6* ± 0.0	73.5* ± 0.1	75.3 ± 1.2	73.6* ± 1.1	73.9 ± 0.4
• Clevr-Count	37.0* ± 0.9	60.7* ± 6.9	56.0* ± 2.9	69.5 ± 4.0	56.3* ± 1.9	59.9* ± 2.6	56.7* ± 1.2	64.9 ± 2.2
• Clevr-Dist	34.4* ± 1.2	56.7* ± 3.6	52.5* ± 3.0	60.3 ± 0.6	59.0* ± 1.5	57.3* ± 1.0	57.5* ± 0.9	60.3 ± 1.0
• DMLab	37.7* ± 0.5	37.2* ± 1.3	47.0 ± 1.4	44.5* ± 0.3	50.0 ± 1.6	44.9 ± 1.2	42.9* ± 0.8	45.2 ± 1.1
• dSpr-Loc	12.8* ± 5.0	65.0 ± 5.3	40.4* ± 8.3	69.3 ± 8.2	76.5 ± 1.8	73.6 ± 2.0	62.8* ± 3.1	64.1 ± 2.4
• dSpr-Ori	28.7* ± 1.1	31.2* ± 9.0	45.0 ± 0.8	49.5 ± 0.8	49.0 ± 1.2	48.7 ± 0.2	41.2* ± 0.5	44.8 ± 3.0
• KITTI-Dist	59.3* ± 1.1	68.2* ± 2.2	82.4 ± 0.5	76.5 ± 2.9	82.7 ± 1.0	82.0 ± 0.6	79.5 ± 0.5	78.0 ± 2.9
• sNORB-Azim	14.1* ± 0.4	11.7* ± 1.3	16.5* ± 0.9	16.7 ± 1.7	18.7 ± 0.7	17.7 ± 0.7	14.9* ± 1.9	16.9 ± 1.0
• sNORB-Elev	23.5* ± 0.1	20.7* ± 1.4	26.9* ± 0.1	33.8 ± 1.8	31.1* ± 1.4	29.3* ± 1.5	29.1* ± 1.0	32.1 ± 1.6

Table 7. **Detailed VTAB-1k benchmark results using a ViT-S/14**. Average and standard deviation accuracies across three seeds for seven adapters across all 19 VTAB-1k datasets. Results marked with * indicate accuracies significantly lower ($p = 0.05$) than the accuracy of PVeRA.

	Linear	Bottleneck	IA^3	AdaptFormer	DoRA	LoRA	VeRA	PVeRA
• Caltech101	84.9* ± 0.5	88.2 ± 0.5	85.0* ± 0.4	86.4* ± 1.3	88.1 ± 0.8	89.0 ± 1.2	89.9 ± 0.9	88.5 ± 0.8
• CIFAR-100	70.3* ± 0.1	50.9* ± 35.3	77.1 ± 0.1	39.3* ± 25.5	76.0* ± 0.9	74.0 ± 0.5	66.4* ± 1.3	75.6 ± 1.4
• DTD	74.0* ± 0.2	77.6 ± 0.4	76.6 ± 0.3	74.8* ± 0.4	79.0 ± 0.1	78.0 ± 0.2	75.4* ± 0.5	77.1 ± 1.1
• Flowers102	99.1* ± 0.2	99.7 ± 0.0	99.7 ± 0.0	66.2* ± 46.3	99.7 ± 0.0	99.7 ± 0.0	99.7 ± 0.0	99.7 ± 0.0
• Pets	93.4 ± 0.9	94.4 ± 0.2	94.2 ± 1.1	86.7* ± 2.0	93.6 ± 1.2	93.3 ± 0.6	92.6* ± 0.8	93.3 ± 1.2
• Sun397	55.7* ± 0.4	56.5* ± 0.4	57.7* ± 0.1	3.4* ± 0.5	57.1* ± 0.2	55.0* ± 0.4	54.0* ± 1.2	58.1 ± 0.1
• SVHN	32.9* ± 0.7	67.4* ± 33.8	44.9* ± 1.7	30.1* ± 9.5	81.1 ± 3.8	92.4* ± 0.7	89.2 ± 0.6	88.9 ± 1.1
• Camelyon	82.7* ± 0.5	86.5 ± 0.4	81.5* ± 0.4	85.8* ± 1.8	84.9 ± 0.6	87.4* ± 1.3	85.7* ± 1.2	86.6 ± 1.0
• EuroSAT	92.0* ± 0.1	94.7* ± 0.3	94.7* ± 0.2	95.1 ± 0.3	95.9* ± 0.3	95.0 ± 0.1	95.9 ± 0.2	95.4 ± 0.2
• Resisc45	80.4* ± 0.8	88.9* ± 1.7	87.7* ± 0.3	81.1* ± 4.9	90.0 ± 0.4	91.3 ± 0.3	88.3* ± 0.7	89.8 ± 0.2
• Retinopathy	73.6 ± 0.0	73.6 ± 0.0	73.6 ± 0.0	73.6 ± 0.0	75.0 ± 1.0	73.6 ± 0.0	73.6 ± 0.0	73.6 ± 0.0
• Clevr-Count	43.4* ± 1.5	72.5* ± 6.6	55.2* ± 1.5	39.6* ± 13.1	85.8* ± 4.5	66.5 ± 3.5	72.1* ± 6.0	82.9 ± 4.2
• Clevr-Dist	31.9* ± 0.4	60.4* ± 1.9	52.3* ± 1.2	41.4* ± 5.6	57.5* ± 0.6	58.9* ± 1.5	60.3* ± 0.7	62.7 ± 0.9
• DMLab	41.4* ± 0.7	53.0 ± 2.5	46.6* ± 0.6	33.7* ± 2.5	53.6 ± 0.5	56.5 ± 0.8	53.2 ± 0.8	52.5 ± 0.4
• dSpr-Loc	9.6* ± 0.6	79.9 ± 6.7	31.7* ± 1.2	17.0* ± 5.2	75.5* ± 3.0	43.0 ± 7.2	75.5 ± 4.0	70.6 ± 7.0
• dSpr-Ori	25.5* ± 2.9	53.1 ± 1.3	47.6* ± 3.0	18.3* ± 1.9	55.8 ± 1.3	53.4 ± 1.3	52.3* ± 1.2	53.0 ± 2.1
• KITTI-Dist	48.7* ± 3.4	82.5* ± 1.7	67.7* ± 1.0	70.8* ± 2.9	83.8 ± 1.2	83.7 ± 1.5	82.3* ± 1.1	84.3 ± 1.2
• sNORB-Azim	12.6* ± 0.1	21.6* ± 3.8	12.6* ± 0.6	7.7* ± 0.6	21.4 ± 0.3	26.2* ± 0.6	24.0 ± 0.7	24.2 ± 0.4
• sNORB-Elev	23.6* ± 0.9	32.3* ± 2.6	27.6* ± 0.0	23.1* ± 1.7	34.4* ± 0.2	35.8* ± 1.6	36.4 ± 1.7	36.6 ± 2.2

Table 8. **Detailed VTAB-1k benchmark results using a ViT-L/14**. Average and standard deviation accuracies across three seeds for seven adapters across all 19 VTAB-1k datasets. Results marked with * indicate accuracies significantly lower ($p = 0.05$) than the accuracy of PVeRA.