# 7. Details for Fine-tuning ConvNeXt-V2-B

Table 8 provides a set of custom hyperparameters used in all experiments. For additional default parameters that are not included in Table 8, please refer to the MMDetection repositories.

We mainly adopt the VTAB-1K Benchmark for image classification. VTAB-1K [67] consists of 19 image classification tasks that cover a broad range of domains, divided into three categories: Natural, Specialized, and Structured. These tasks encompass a wide array of potential downstream applications, making the benchmark a robust indicator of a method's transfer learning abilities. Each dataset is composed of 800 training samples and 200 validation samples. In line with previous work [25, 26], we fine-tune the pre-trained ConvNeXt-V2-B model using all 1,000 training and validation samples and evaluate its performance on the test set. Consistent with [25, 30], we use unnormalized inputs, as done in the original VTAB paper [67].

# 8. Details for Generation Tasks

We fine-tune text-to-image diffusion model, SDXL5 [39], designed for subject-specific generation tasks, as presented in recent work [45]. The goal of this task is to generate images that closely align with prompts associated with a specific subject, defined by a few example images. This process begins by fine-tuning a text-to-image model using image-text pairs, where the text includes a unique identifier (e.g., "A picture of a [V] cat"). Afterward, the model generates images based on new prompts that include this identifier, with the aim of creating images that reflect the learned subject. Fine-tuning is performed with a learning rate of 1e-4 and a batch size of 4. The model is trained for 500 steps on a single 80GB A100 GPU, taking approximately 21 minutes to finish. During the generation phase, we run 50 inference steps per prompt to produce the final images, which takes about 30 seconds. We primarily use the official DreamBooth dataset [45] for the diffusion process.

# 9. Details for Commonsense Reasoning Tasks

The commonsense reasoning benchmarks comprise 8 different sub-tasks, each with a distinct dataset, including BoolQ [12], PIQA [6], SIQA [47], HellaS. [66], WinoG. [46], ARC-e/ARC-c [13], and OBQA [36]. In accordance with the protocol described by [24], we combine the training datasets from all sub-tasks to form the Commonsense170K dataset, and perform evaluations on each subtask's respective testing set.

We incorporate results from ChatGPT's implementation using the gpt-3.5-turbo API, specifically focusing on zero-shot Chain of Thought approaches [62]. For fair comparison, the initial fine-tuning for models utilizing LieRA is performed under LoRA configurations with the learning rate being the only variable optimized for better performance. The hyper-parameter settings for LieRA are shown in Table 9. The results for LoRA are taken from [24, 33].

# 10. Details for NLU Tasks

For the natural language understanding (NLU) task, we use the General Language Understanding Evaluation (GLUE) benchmark [58], which is designed to assess performance across a variety of tasks. The benchmark includes two single-sentence classification tasks, CoLA [61] and SST-2 [55], three similarity and paraphrase tasks: MRPC [16], QQP [58], and STS-B [7], and three natural language inference tasks: MNLI [63], QNLI [42], and RTE [3, 4, 14, 18]. We fine-tune both the DeBERTaV3-base and DeBERTaV3-large models [21] for this task. The corresponding hyper-parameter settings are listed in Table 10.

Table 8. Hyper-parameter setup for object detection and segmentation.

| **Dataset** | **Toolkit** | **Model** | **Schedule** | **LR** | **BS** | **Optimizer** | **Weight Decay** |
|---|---|---|---|---|---|---|---|
| COCO | MMDetection [8] | Mask R-CNN [20] | 12ep | 1e-4 | 32 | AdamW | 5e-2 |

Table 9. Hyper-parameter settings on commonsense reasoning tasks.

| **Hyper-parameters** | LLaMA-7B | | LLaMA3-8B | |
|---|---|---|---|---|
| Rank $r$ | 16 | 32 | 16 | 32 |
| $\alpha$ | 32 | 64 | 32 | 64 |
| LR | 2e-3 | 1e-3 | 6e-4 | 8e-4 |
| LR Scheduler | Linear | | | |
| Dropout | 0.05 | | | |
| Optimizer | AdamW | | | |
| Batch size | 16 | | | |
| Warmup Steps | 100 | | | |
| Epochs | 3 | | | |
| Where | Q, K, V, Up, Down | | | |

Table 10. Hyper-parameter settings on NLU task.

| Hyper-parameter | MNLI | SST-2 | CoLA | QQP | QNLI | RTE | MRPC | STS-B |
|---|---|---|---|---|---|---|---|---|
| Optimizer | AdamW | | | | | | | |
| Warmup Ratio | 0.1 | | | | | | | |
| LR schedule | Linear | | | | | | | |
| Rank $r$ | 2 & 8 | | | | | | | |
| LoRA alpha | 4 & 16 | | | | | | | |
| Max Seq. Len. | 256 | 128 | 64 | 320 | 512 | 320 | 320 | 128 |
| Batch Size | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 |
| Learning Rate | 5e-3 | 8e-3 | 8e-3 | 1e-3 | 5e-3 | 2e-3 | 1e-3 | 5e-3 |
| Epochs | 12 | 24 | 25 | 5 | 5 | 50 | 30 | 25 |