### **Overview of Datasets and Architectures**

* **Datasets:**
  + **DISC:** Refers to models trained on the DISC dataset.
  + **IMAGENET:** Refers to models trained on ImageNet.
* **Architecture Implementations:**
  + **CLASSY:** Models implemented with Classy Vision (Facebook’s research framework).
  + **TORCHVISION:** Models that follow the standard torchvision API.
  + **TORCHSCRIPT:** Models that have been exported using TorchScript for optimized deployment.

### **What Do the Modifiers Mean?**

Each of the following modifiers—**MIXUP**, **BLUR**, **ADVANCED**, and **LARGE**—indicates specific modifications or augmentations applied during training, or certain design choices in the model:

#### **1. MIXUP**

* **Definition:** Mixup is a data augmentation strategy that constructs new training examples by taking a **linear interpolation** of two inputs and their corresponding labels.
* **Purpose:** This blending of examples helps to regularize the model and smooth its decision boundaries, which can lead to better generalization and improved robustness against adversarial examples.
* **Application in SSCD Models:** Models with the "MIXUP" label have been trained with this augmentation technique, which might make them better at handling variations in the input data.

#### **2. BLUR**

* **Definition:** The term "BLUR" generally indicates the use of a Gaussian blur (or another blurring technique) as part of the data augmentation process.
* **Purpose:** Blurring images during training forces the network to focus on larger-scale, more robust features rather than high-frequency details. This can make the model more robust to variations in focus or slight misalignments in real-world scenarios.
* **Application in SSCD Models:** Models marked with "BLUR" likely experienced training scenarios where images were intentionally blurred to challenge the network to learn more invariant features.

#### **3. ADVANCED**

* **Definition:** The "ADVANCED" designation generally implies that the model was trained using a more sophisticated augmentation or training regimen.
* **Purpose:** This might include a combination of multiple augmentation techniques such as mixup, blur, additional color variations, random crops, or even advanced loss functions and optimization strategies. The goal is to further improve the model’s robustness and accuracy.
* **Application in SSCD Models:** SSCD models with "ADVANCED" in their name would typically employ a state-of-the-art training pipeline that leverages several augmentation methods together to enhance overall performance.

#### **4. LARGE**

* **Definition:** "LARGE" denotes that the model uses a larger backbone or more parameters compared to its standard counterpart.
* **Purpose:** By increasing the capacity (i.e., more layers or wider layers) the model can potentially achieve better performance on complex data. However, this also means higher computational cost during training and inference.
* **Application in SSCD Models:** A "LARGE" model might be chosen for scenarios where the extra computational expense is acceptable in exchange for improved accuracy and richer feature representations.

### **Practical Considerations**

* **Performance and Robustness:** Although each variant uses the same base weight (for example, sscd\_disc\_mixup), the differences in augmentation and model capacity can impact how the network performs on test data. In practice:  
  + **Mixup models** might offer smoother decision boundaries.
  + **Blur models** may be more robust to noisy or blurred inputs.
  + **Advanced models** typically incorporate multiple techniques to boost overall performance.
  + **Large models** could achieve higher accuracy due to increased capacity but might be slower in inference.
* **Deployment and Inference:** While the underlying weight parameters can be the same, the way the network processes input might slightly differ based on the architecture implementation (e.g., CLASSY vs. TORCHVISION vs. TORCHSCRIPT). These differences are generally minor if the same augmentation or model modifications are applied during training, but might affect performance metrics (like latency) and numerical precision.

### **Summary Table**

| **Modifier** | **Training Modification / Design Choice** | **Expected Impact** |
| --- | --- | --- |
| **MIXUP** | Applies mixup data augmentation (combining pairs of inputs/labels) | Better regularization, smoother decision boundaries |
| **BLUR** | Incorporates Gaussian (or similar) blur as augmentation | Improved robustness against noisy or out-of-focus inputs |
| **ADVANCED** | Uses a combination of several advanced techniques (potentially mixup, blur, etc.) | Enhanced overall performance and robustness |
| **LARGE** | Utilizes a larger, higher capacity backbone | Higher accuracy (at the cost of increased compute) |

This detailed explanation should help clarify the nuances behind the naming of your SSCD models. Each modifier addresses a specific goal in improving the network’s performance, either through training augmentation techniques or architectural modifications.