Appendices

A. BiasBed

BiasBed is a Python package that can simply be installed with pip package manager. Once installed, we can run sets of experiments with a single command. We follow by default a common training protocol: the learning rate is set to $1e$-4, the learning decay after 30, 60 and 90 epochs is 0.1, the SGD optimizer has momentum 0.9 with weight decay $1e$-4, and we train for 100 epochs in total. We use ResNet-50 as a backbone model for all experiments.

A.1. Adding algorithms

New algorithms in *BiasBed* can easily be added by extending the provided abstract algorithm class. The important function to implement here is `update(x, y)`. This update function receives batches of input data and corresponding ground truth. It needs to predict the logits, compute the loss and backpropagate the gradients. To implement empirical risk minimization [40], for example, we add a folder `ERM` in the algorithms folder and add `algorithm.py` with the corresponding `update(x, y)` function:

```
def update(self, x, y):
    self.optimizer.zero_grad(set_to_none=True)
    with autocast(enabled=self.algorithm_cfg.mixedprec):
        logits = self.model(x)
        loss = self.loss(logits, y)

    # Backward loss
    self.scaler.scale(loss).backward()
    self.scaler.step(self.optimizer)
    self.scaler.update()
```

Each algorithm folder contains two additional config files `config.yaml` and `sweep.yaml`. The former includes all algorithm specific (hyper-) parameters:

```
# Include default parameters of your algorithm here.
mixedprec: True
backbone: resnet50 # Net for MNIST
optimizer: SGD
learning_rate: 1e-4
milestones:
  - 30
  - 60
  - 90
momentum: 0.9
weight_decay: 1e-4
gamma: 0.1
```

and the latter includes all (hyper-) parameters necessary for sweeping over individual parameters in the algorithm and main config files, *e.g.*

```
# Include sweep parameters of your algorithm here
parameters:
  epoch:
    values:
      - 10
```

```
      - 20
      - 30
  momentum:
    values:
      - 0.7
      - 0.8
      - 0.9
```

BiasBed takes care of registering and adding the algorithm to the framework. See Appendix A.3 on how to run experiments with the newly added algorithm.

A.2. Adding datasets

Adding dataloaders for datasets is equally simple. We need to add a folder with the dataloader name and implement the `train_loader, val_loader, test_loader` from the dataloader template in a file called `dataloader.py`, *e.g.* in the case of Cue-Conflict

```
def train_loader(self) -> Iterable:
    data_loader = DataLoader(self.dataset,
                             batch_size=self.config.training.batch_size,
                             num_workers=self.config.num_workers,
                             sampler=sampler)
    return data_loader
```

where `self.dataset` is a PyTorch `ImageFolder` dataset. Of course, each dataset comes with its own config file to include dataset specific (hyper-) parameters. Once the dataloader is added, the dataset can be seamlessly added to the main config file detailed in the following section.

A.3. Running BiasBed

BiasBed supports various training modes, including full support of half precision, multi-GPU or hyperparameter sweeping with cluster support. We provide a fully flexible main configuration file to activate or deactivate settings. Only a single line has to be edited for training an algorithm on a single GPU or on a compute node with multiple GPUs. We fully integrated Weights and Biases [3] into our framework, too, such that these are automatically used to search and tune hyperparameters. A *BiasBed* user can either use our code to launch runs on common high performance computing environments or easily add a custom launcher. We can start a single run with the command:

```
biasedbed single # Starts a single run with the algorithm set in config.yaml
```

or start automated hyperparameter tuning with

```
biasedbed sweep # Starts a sweep with settings from sweeping/sweep.yaml
```

In both cases, algorithm and dataset settings can be edited and activated in `config.yaml`. For example, we can train SagNet [31] with eight GPUs on standard ImageNet and evaluate the model on six different datasets with the following configuration:

```
# Distributed training
num_workers: 8
distributed: 1
world_size: 8
# Algorithm
```

```yaml
algorithm:
  name: SagNet

# Training
training:
  dataset:
    name: ImageNet1k
  epochs: 100

# Testing
testing:
  datasets:
    - CueConflict
    - Silhouette
    - Sketch
    - Edge
    - ImageNetStylized
    - ImageNet1k
  interval: 1
```

B. Algorithms

ERM "Empirical Risk Minimization" [40] minimizes the cross entropy loss across the training data and serves as our baseline algorithm.

Stylized ImageNet "Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness" [13] is the first paper to recognize and rigorously demonstrate texture bias in existing neural architectures. To reduce texture bias, the authors propose a stylized version of ImageNet, where they use AdaIN [18] to change the texture of one image with another random image of ImageNet.

Debiased "Shape-texture debiased neural network training" [27] extends the idea of [13] by augmenting the dataset online, *i.e*. when feeding a batch of (original ImageNet) images into the network. Instead of only training on the content label, a convex combination of the style image class label and the content class label is used to guide the network to "debiased" weights, *i.e*. the network is forced to predict the content class solely from shape cues and the style class solely from texture cues. The authors argue, that performance is generally higher on all tested datasets (and not only on shape-biased sets) compared to [13].

DeepAugment "The Many Faces of Robustness: A Critical Analysis of Out-of-Distribution Generalization" [15] introduces additional deep augmentation techniques similar to [13]. In DeepAugment, an image is passed through an image-to-image network, but the forward pass is distorted by an altering the network. This distorts the resulting image in a similar way to augmentation methods. The authors defined a number of pertubations such as zeroing, negating, convolving, transposing, or switching activation functions and drew per image random samples from them. The networks used in DeepAugment are the pre-trained networks EDSR by [28] and CAE [39]. The resulting augmented images for ImageNet are provided at https://github.com/hendrycks/imagenet-r/tree/master/DeepAugment. In principle, they can be combined with any other algorithm by appending them to the standard ImageNet dataset.

InfoDrop "Informative Dropout for Robust Representation Learning: A Shape-bias Perspective" [37] proposes an agnostic light-weight method to reduce texture bias in neural networks. The main idea is to enforce visual primitives such as edges and corners (*i.e*., regions with high shape information) and to reduce homogeneous and repetitive patterns (*i.e*., regions with low shape information). During training, neurons corresponding to input patches with low shape information are more likely to be zeroed out than neurons with high shape information patches.

SagNet "Reducing Domain Gap by Reducing Style Bias" [31] introduces a style-agnostic network that becomes invariant to texture with a style randomization and content randomization network. Features of a shared encoder are randomly interpolated with style features from another image in the batch. The style network is forced to predict the correct style label and the content network needs to predict the correct content label. The gradients of the former network are adversarially used to update the shared encoder.

pAdaIN "Permuted AdaIN: Reducing the Bias Towards Global Statistics in Image Classification" [34] follows a similar idea as SagNet but only incorporates a single style network that is forced to predict the correct content label from style-interpolated features.

C. Full BiasBed results

We report all individual results per algorithm according to the best in-domain validation score and average score over the last 30 epochs.

C.1. Model selection: average of last 30 epochs

C.1.1 ERM [40]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	73.2 ± 0.3	47.2 ± 1.4	23.9 ± 1.7	56.4 ± 0.9	22.0 ± 0.6	7.6 ± 0.2	2.0 ± 0.1	21.8 ± 0.5	43.5 ± 0.6	47.6 ± 0.5
2	73.1 ± 0.3	45.4 ± 2.4	28.9 ± 2.5	56.7 ± 0.8	22.8 ± 0.7	7.6 ± 0.3	1.7 ± 0.1	22.6 ± 0.4	43.7 ± 0.5	47.5 ± 0.5
3	73.6 ± 0.3	45.6 ± 2.5	19.4 ± 1.8	55.4 ± 1.0	22.6 ± 0.7	7.5 ± 0.2	2.1 ± 0.2	22.8 ± 0.4	43.8 ± 0.5	48.0 ± 0.5
4	73.2 ± 0.3	47.6 ± 2.0	17.9 ± 2.0	55.5 ± 1.1	21.6 ± 0.7	8.0 ± 0.3	2.2 ± 0.2	22.3 ± 0.4	43.8 ± 0.5	47.5 ± 0.4
5	73.5 ± 0.3	46.3 ± 1.3	24.5 ± 2.8	56.4 ± 1.0	22.1 ± 0.8	7.7 ± 0.3	2.1 ± 0.1	22.6 ± 0.4	44.1 ± 0.6	48.7 ± 0.4
6	73.2 ± 0.3	47.1 ± 2.2	21.0 ± 2.4	55.6 ± 1.3	21.3 ± 0.4	7.7 ± 0.3	2.2 ± 0.2	22.8 ± 0.3	43.8 ± 0.5	47.8 ± 0.4
7	73.3 ± 0.3	46.1 ± 2.5	23.6 ± 2.3	57.5 ± 1.1	21.4 ± 0.8	7.8 ± 0.2	2.0 ± 0.2	22.5 ± 0.3	43.6 ± 0.6	48.0 ± 0.5
8	73.3 ± 0.4	48.5 ± 1.8	21.0 ± 2.4	56.6 ± 1.2	23.1 ± 0.8	7.6 ± 0.2	2.1 ± 0.2	22.5 ± 0.5	44.0 ± 0.5	47.4 ± 0.6
9	73.5 ± 0.3	48.9 ± 1.5	21.7 ± 2.1	56.2 ± 1.0	22.2 ± 0.6	7.5 ± 0.2	2.1 ± 0.1	22.3 ± 0.3	44.1 ± 0.5	47.5 ± 0.4
10	73.5 ± 0.3	45.4 ± 2.6	23.7 ± 2.4	57.1 ± 1.3	22.4 ± 0.6	7.5 ± 0.2	2.0 ± 0.1	22.7 ± 0.4	44.0 ± 0.4	48.1 ± 0.4

C.1.2 Debiased [27]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	74.0 ± 0.3	49.3 ± 3.0	32.7 ± 2.6	59.8 ± 0.8	29.4 ± 0.5	15.6 ± 0.4	2.6 ± 0.2	27.5 ± 0.4	49.3 ± 0.3	51.0 ± 0.5
2	73.9 ± 0.3	48.7 ± 1.9	20.1 ± 1.3	61.1 ± 1.1	28.7 ± 0.8	15.4 ± 0.5	2.4 ± 0.2	27.0 ± 0.6	49.0 ± 0.6	51.0 ± 0.5
3	74.1 ± 0.3	48.0 ± 1.8	36.3 ± 2.0	60.2 ± 1.0	29.6 ± 0.8	15.8 ± 0.5	2.5 ± 0.2	27.4 ± 0.5	49.0 ± 0.5	51.2 ± 0.4
4	74.0 ± 0.3	46.1 ± 1.9	27.3 ± 2.3	60.8 ± 1.0	30.3 ± 0.5	15.9 ± 0.4	2.5 ± 0.2	27.0 ± 0.4	49.1 ± 0.5	50.5 ± 0.5
5	73.7 ± 0.3	47.6 ± 1.6	32.7 ± 3.0	59.3 ± 1.3	29.1 ± 0.5	15.3 ± 0.4	2.5 ± 0.1	26.9 ± 0.4	48.9 ± 0.6	50.9 ± 0.5
6	74.1 ± 0.4	46.9 ± 2.0	27.7 ± 2.6	60.6 ± 0.9	29.6 ± 0.7	15.8 ± 0.5	2.7 ± 0.2	27.7 ± 0.4	49.6 ± 0.5	51.1 ± 0.5
7	74.0 ± 0.4	44.5 ± 2.3	31.2 ± 2.4	59.0 ± 1.1	27.1 ± 0.5	15.4 ± 0.5	2.5 ± 0.2	27.1 ± 0.3	49.0 ± 0.5	50.8 ± 0.5
8	73.9 ± 0.3	50.5 ± 1.5	31.6 ± 3.3	60.0 ± 1.1	28.9 ± 0.8	15.9 ± 0.5	2.7 ± 0.1	27.0 ± 0.3	49.0 ± 0.5	51.0 ± 0.5
9	73.9 ± 0.3	50.2 ± 2.0	27.1 ± 2.4	61.3 ± 0.9	30.2 ± 0.6	15.7 ± 0.4	2.5 ± 0.2	27.4 ± 0.5	49.2 ± 0.5	50.6 ± 0.6
10	74.0 ± 0.3	51.2 ± 1.2	27.3 ± 1.9	59.4 ± 0.8	28.0 ± 0.5	15.4 ± 0.5	2.7 ± 0.2	27.1 ± 0.4	49.1 ± 0.6	50.3 ± 0.6

C.1.3 DeepAug ERM (CAE) [15]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	72.9 ± 0.3	52.5 ± 1.3	32.8 ± 2.4	62.5 ± 1.1	30.9 ± 0.7	12.9 ± 0.3	2.8 ± 0.1	27.9 ± 0.5	62.6 ± 0.4	56.2 ± 0.5
2	72.9 ± 0.3	51.1 ± 1.4	34.8 ± 1.8	64.0 ± 1.1	30.6 ± 0.6	13.0 ± 0.2	2.8 ± 0.1	27.7 ± 0.4	62.6 ± 0.3	55.8 ± 0.4
3	72.5 ± 0.3	51.3 ± 1.2	39.2 ± 2.9	64.4 ± 1.3	29.4 ± 0.7	13.0 ± 0.3	2.6 ± 0.1	27.5 ± 0.4	62.3 ± 0.4	55.8 ± 0.4
4	73.0 ± 0.2	50.5 ± 1.7	29.1 ± 2.0	65.0 ± 0.7	30.3 ± 0.7	12.9 ± 0.2	2.9 ± 0.1	27.6 ± 0.4	62.6 ± 0.3	56.0 ± 0.4
5	72.8 ± 0.3	51.0 ± 1.7	32.1 ± 3.1	65.0 ± 0.9	29.3 ± 0.8	13.0 ± 0.2	2.6 ± 0.1	27.3 ± 0.4	62.4 ± 0.4	56.2 ± 0.4
6	73.0 ± 0.3	53.4 ± 1.9	39.6 ± 1.4	63.3 ± 1.0	30.6 ± 0.8	12.8 ± 0.2	2.7 ± 0.1	28.1 ± 0.4	62.7 ± 0.4	56.5 ± 0.3
7	72.7 ± 0.3	53.7 ± 1.6	36.6 ± 2.2	62.8 ± 1.0	30.9 ± 0.7	13.2 ± 0.3	2.6 ± 0.1	27.5 ± 0.4	62.5 ± 0.3	56.4 ± 0.4
8	72.9 ± 0.3	50.3 ± 1.4	32.2 ± 2.3	61.4 ± 1.0	29.2 ± 1.0	12.9 ± 0.3	2.8 ± 0.2	27.4 ± 0.4	62.6 ± 0.4	56.4 ± 0.3
9	72.7 ± 0.3	48.9 ± 1.9	39.4 ± 2.5	62.5 ± 1.2	29.9 ± 0.7	13.1 ± 0.3	2.8 ± 0.1	27.1 ± 0.4	62.2 ± 0.4	55.7 ± 0.4

C.1.4 DeepAug ERM (EDSR) [15]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	71.8 ± 0.7	52.3 ± 2.2	32.2 ± 2.9	61.5 ± 1.5	31.5 ± 1.0	10.5 ± 0.4	2.0 ± 0.2	25.4 ± 0.6	51.1 ± 0.9	64.0 ± 0.8
2	71.8 ± 0.7	51.9 ± 1.7	35.0 ± 3.7	61.6 ± 1.8	33.2 ± 1.0	10.4 ± 0.5	2.0 ± 0.1	25.8 ± 0.5	51.0 ± 0.8	64.1 ± 0.8
3	71.9 ± 0.6	51.0 ± 1.9	30.9 ± 2.2	60.3 ± 1.6	31.1 ± 0.9	10.9 ± 0.5	1.9 ± 0.2	25.8 ± 0.5	51.4 ± 0.9	64.1 ± 0.8
4	71.9 ± 0.7	52.4 ± 2.1	31.7 ± 3.0	61.6 ± 1.6	32.0 ± 1.1	10.7 ± 0.5	2.1 ± 0.2	25.8 ± 0.7	51.5 ± 0.8	64.2 ± 0.8
5	72.1 ± 0.6	52.2 ± 2.1	37.4 ± 3.6	60.7 ± 1.4	33.3 ± 0.8	10.4 ± 0.5	2.1 ± 0.2	25.9 ± 0.5	51.2 ± 0.9	64.5 ± 0.8
6	72.1 ± 0.6	51.4 ± 1.6	32.1 ± 3.5	59.7 ± 1.6	32.4 ± 1.0	10.7 ± 0.4	2.1 ± 0.2	25.9 ± 0.6	51.1 ± 0.8	64.2 ± 0.8
7	71.7 ± 0.6	51.3 ± 2.4	29.3 ± 3.0	58.6 ± 1.0	33.3 ± 0.9	10.4 ± 0.4	2.1 ± 0.2	25.9 ± 0.5	51.1 ± 0.9	64.0 ± 0.8
8	71.9 ± 0.7	49.9 ± 2.6	29.5 ± 3.2	59.5 ± 1.6	31.8 ± 0.9	10.7 ± 0.5	1.8 ± 0.2	25.1 ± 0.6	50.7 ± 0.8	64.2 ± 0.8
9	71.8 ± 0.7	51.7 ± 2.0	27.5 ± 3.0	59.0 ± 1.4	31.8 ± 0.9	11.0 ± 0.5	1.9 ± 0.1	25.8 ± 0.7	51.2 ± 0.8	64.1 ± 0.8
10	72.0 ± 0.7	49.8 ± 2.2	34.8 ± 3.2	61.1 ± 1.7	32.3 ± 0.9	10.6 ± 0.5	2.3 ± 0.2	26.1 ± 0.4	51.1 ± 0.8	64.2 ± 0.8

C.1.5 Stylized ImageNet [13]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	56.2 ± 0.4	47.7 ± 1.3	61.8 ± 1.9	71.1 ± 0.7	54.5 ± 1.0	52.6 ± 0.5	0.8 ± 0.1	25.5 ± 0.4	39.8 ± 0.4	40.8 ± 0.4
2	55.3 ± 0.5	46.7 ± 1.7	59.3 ± 2.8	70.1 ± 0.8	54.0 ± 1.0	52.8 ± 0.5	0.7 ± 0.1	24.7 ± 0.4	38.5 ± 0.6	40.0 ± 0.5
3	54.6 ± 0.5	43.7 ± 2.9	61.6 ± 2.6	69.5 ± 1.0	55.7 ± 1.1	52.1 ± 0.7	0.7 ± 0.1	24.7 ± 0.4	38.4 ± 0.5	39.6 ± 0.4
4	55.4 ± 0.5	47.8 ± 1.7	61.4 ± 2.8	69.6 ± 1.1	54.0 ± 1.2	52.4 ± 0.6	0.8 ± 0.1	25.1 ± 0.4	38.8 ± 0.5	40.2 ± 0.4
5	55.5 ± 0.5	46.7 ± 1.4	58.4 ± 2.9	71.1 ± 0.8	52.5 ± 1.4	53.0 ± 0.5	0.7 ± 0.1	25.3 ± 0.4	38.7 ± 0.4	40.3 ± 0.4
6	54.8 ± 0.5	50.8 ± 1.4	59.9 ± 2.7	69.1 ± 0.9	54.0 ± 1.4	52.1 ± 0.7	0.7 ± 0.1	24.6 ± 0.4	38.3 ± 0.6	39.5 ± 0.5
7	55.2 ± 0.5	45.1 ± 2.2	56.0 ± 2.5	70.6 ± 1.0	54.1 ± 0.8	52.3 ± 0.6	0.8 ± 0.1	24.7 ± 0.5	38.3 ± 0.6	40.0 ± 0.5
8	55.9 ± 0.4	46.6 ± 1.8	56.3 ± 2.9	71.0 ± 0.6	54.2 ± 1.3	52.6 ± 0.5	0.8 ± 0.1	25.5 ± 0.5	39.1 ± 0.5	40.6 ± 0.5
9	55.3 ± 0.6	49.2 ± 1.8	62.0 ± 3.1	68.7 ± 1.0	55.7 ± 1.2	52.4 ± 0.7	0.7 ± 0.1	25.1 ± 0.3	38.8 ± 0.6	40.1 ± 0.5
10	54.9 ± 0.4	48.3 ± 2.0	58.2 ± 3.5	70.7 ± 0.9	53.3 ± 1.4	52.0 ± 0.5	0.7 ± 0.1	24.7 ± 0.3	38.8 ± 0.4	40.0 ± 0.4

C.1.6 InfoDrop [37]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	73.2 ± 0.4	43.9 ± 1.7	22.4 ± 3.0	57.9 ± 1.1	23.1 ± 0.7	7.8 ± 0.3	2.3 ± 0.2	23.2 ± 0.5	44.3 ± 0.5	48.1 ± 0.4
2	72.1 ± 0.7	47.4 ± 2.0	22.9 ± 3.1	57.1 ± 1.7	21.9 ± 0.7	7.5 ± 0.4	2.0 ± 0.2	22.3 ± 0.6	43.0 ± 0.9	48.0 ± 0.7
3	72.8 ± 0.3	47.1 ± 2.0	22.1 ± 1.5	55.7 ± 1.1	22.4 ± 0.5	7.6 ± 0.3	2.1 ± 0.2	22.6 ± 0.3	44.0 ± 0.5	48.6 ± 0.3
4	73.0 ± 0.4	48.0 ± 2.4	15.5 ± 1.4	55.9 ± 1.2	22.5 ± 0.6	7.6 ± 0.2	2.1 ± 0.1	22.6 ± 0.4	44.0 ± 0.5	48.9 ± 0.4
5	72.8 ± 0.3	46.3 ± 1.4	18.2 ± 1.9	56.4 ± 0.9	22.9 ± 0.7	7.6 ± 0.3	2.2 ± 0.2	22.6 ± 0.5	43.7 ± 0.5	48.2 ± 0.4
6	73.1 ± 0.4	46.5 ± 1.9	16.8 ± 2.8	58.5 ± 1.5	22.9 ± 0.6	7.6 ± 0.3	2.2 ± 0.2	22.9 ± 0.4	44.2 ± 0.5	49.0 ± 0.5
7	72.9 ± 0.3	49.7 ± 1.2	13.2 ± 1.4	55.1 ± 1.2	23.7 ± 0.6	7.8 ± 0.2	2.2 ± 0.2	22.5 ± 0.4	44.0 ± 0.5	48.6 ± 0.5
8	73.1 ± 0.4	47.4 ± 1.6	23.8 ± 2.2	55.8 ± 0.9	22.8 ± 0.8	7.9 ± 0.2	2.1 ± 0.2	22.7 ± 0.4	44.0 ± 0.6	48.2 ± 0.5
9	72.8 ± 0.3	50.1 ± 1.6	13.1 ± 1.3	55.2 ± 1.3	23.7 ± 0.6	7.8 ± 0.2	2.2 ± 0.1	22.6 ± 0.4	44.0 ± 0.5	48.6 ± 0.5
10	72.8 ± 0.3	43.4 ± 1.7	21.5 ± 2.2	58.0 ± 1.0	23.3 ± 0.8	8.2 ± 0.3	2.3 ± 0.1	22.8 ± 0.5	44.2 ± 0.4	47.8 ± 0.3

C.1.7 SagNet [31]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	74.8 ± 0.3	45.2 ± 1.6	29.8 ± 1.9	59.6 ± 0.9	19.8 ± 0.4	6.4 ± 0.2	1.8 ± 0.2	22.8 ± 0.5	44.1 ± 0.5	48.1 ± 0.3
2	73.7 ± 0.2	44.9 ± 0.7	25.4 ± 1.1	60.4 ± 1.2	20.1 ± 0.4	6.1 ± 0.2	1.7 ± 0.1	22.4 ± 0.5	42.9 ± 0.5	47.2 ± 0.3
3	73.7 ± 0.1	44.0 ± 1.1	28.8 ± 2.1	60.3 ± 1.1	20.1 ± 0.4	6.3 ± 0.2	1.7 ± 0.1	22.4 ± 0.4	43.1 ± 0.5	47.1 ± 0.3
4	73.6 ± 0.2	45.0 ± 1.0	27.0 ± 2.3	59.9 ± 0.9	19.3 ± 0.4	6.2 ± 0.2	1.5 ± 0.1	22.4 ± 0.4	43.0 ± 0.5	47.1 ± 0.4
5	73.6 ± 0.1	43.4 ± 1.0	24.2 ± 1.8	60.1 ± 1.2	19.4 ± 0.4	6.2 ± 0.2	1.7 ± 0.1	22.4 ± 0.5	43.1 ± 0.4	47.2 ± 0.3
6	74.7 ± 0.3	45.4 ± 1.1	27.9 ± 1.8	59.7 ± 1.0	19.9 ± 0.5	6.4 ± 0.2	1.8 ± 0.1	22.8 ± 0.4	44.0 ± 0.5	48.1 ± 0.4
7	73.6 ± 0.1	45.9 ± 1.2	27.1 ± 2.5	61.1 ± 1.2	19.6 ± 0.4	6.1 ± 0.2	1.8 ± 0.2	22.5 ± 0.5	43.0 ± 0.5	47.0 ± 0.3
8	73.6 ± 0.2	44.5 ± 1.2	27.2 ± 1.9	60.9 ± 1.0	19.4 ± 0.4	6.1 ± 0.2	1.6 ± 0.1	22.4 ± 0.4	43.0 ± 0.6	47.2 ± 0.4
9	73.6 ± 0.1	43.5 ± 1.2	27.2 ± 1.9	61.0 ± 1.1	20.1 ± 0.4	6.2 ± 0.2	1.5 ± 0.1	22.5 ± 0.4	43.1 ± 0.5	47.2 ± 0.3
10	73.6 ± 0.2	44.6 ± 1.2	26.4 ± 1.7	60.9 ± 1.3	19.4 ± 0.4	6.2 ± 0.2	1.7 ± 0.2	22.4 ± 0.5	42.9 ± 0.6	47.1 ± 0.4

C.1.8 pAdaIN [34]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	72.9 ± 0.1	45.0 ± 0.9	18.8 ± 1.9	55.4 ± 0.7	21.6 ± 0.4	8.3 ± 0.1	1.4 ± 0.1	21.2 ± 0.2	42.7 ± 0.2	49.0 ± 0.2
2	73.0 ± 0.1	47.0 ± 0.8	13.3 ± 0.9	56.1 ± 0.9	21.7 ± 0.4	8.3 ± 0.1	1.4 ± 0.1	21.5 ± 0.2	42.7 ± 0.2	48.1 ± 0.2
3	73.0 ± 0.1	44.5 ± 1.2	25.0 ± 1.0	57.0 ± 0.7	20.5 ± 0.4	8.0 ± 0.1	1.5 ± 0.1	21.2 ± 0.2	42.8 ± 0.2	48.6 ± 0.2
4	73.0 ± 0.1	44.4 ± 1.2	21.7 ± 3.4	56.0 ± 1.1	21.7 ± 0.4	8.3 ± 0.1	1.5 ± 0.1	21.4 ± 0.3	42.7 ± 0.2	48.9 ± 0.2
5	73.0 ± 0.1	40.7 ± 1.4	18.6 ± 1.3	56.6 ± 0.7	20.3 ± 0.3	7.8 ± 0.1	1.5 ± 0.1	21.7 ± 0.2	42.4 ± 0.2	48.3 ± 0.2
6	73.0 ± 0.1	46.7 ± 1.0	23.1 ± 1.1	57.4 ± 0.7	22.1 ± 0.4	8.1 ± 0.1	1.4 ± 0.1	21.4 ± 0.2	42.8 ± 0.2	48.5 ± 0.2
7	73.1 ± 0.1	43.7 ± 1.0	22.0 ± 0.9	55.9 ± 0.6	21.3 ± 0.5	8.1 ± 0.1	1.5 ± 0.1	21.8 ± 0.2	42.7 ± 0.2	48.6 ± 0.2
8	73.2 ± 0.1	45.8 ± 1.1	19.7 ± 0.8	56.9 ± 0.7	20.7 ± 0.4	8.1 ± 0.1	1.5 ± 0.1	21.3 ± 0.2	42.7 ± 0.2	48.6 ± 0.2
9	73.1 ± 0.1	45.0 ± 2.2	23.2 ± 4.2	56.4 ± 0.7	22.0 ± 0.3	8.2 ± 0.1	1.5 ± 0.1	21.7 ± 0.2	42.8 ± 0.2	48.7 ± 0.2
10	73.1 ± 0.1	40.8 ± 1.2	23.4 ± 1.3	56.6 ± 0.5	21.3 ± 0.4	7.9 ± 0.1	1.4 ± 0.1	21.3 ± 0.2	42.3 ± 0.1	48.3 ± 0.1

C.2. Model selection: best validation score

C.2.1 ERM [40]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	73.6	48.1	23.1	56.2	22.0	7.8	2.0	22.2	44.1	47.6
2	73.5	42.5	26.2	56.5	23.0	7.9	1.7	23.5	44.5	48.3
3	74.0	46.9	20.0	54.3	22.0	8.0	1.9	22.6	44.9	48.6
4	73.7	50.0	18.1	56.0	21.9	8.0	2.4	22.8	44.1	47.6
5	73.9	46.3	25.6	56.4	21.3	7.9	2.2	22.9	44.6	49.1
6	73.7	48.1	20.6	56.5	22.7	7.8	2.0	23.2	44.3	48.0
7	73.8	48.1	22.5	57.6	20.9	8.0	2.3	22.5	44.0	48.3
8	73.8	50.0	20.0	56.0	24.2	7.8	2.0	23.2	44.7	48.1
9	73.9	48.8	21.3	54.3	21.7	8.0	2.1	22.6	44.7	48.8
10	73.9	44.4	28.7	55.9	22.0	7.5	2.0	22.6	44.4	48.4

C.2.2 Debiased [27]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	74.4	51.9	33.1	60.1	28.8	15.9	2.7	27.4	49.5	51.7
2	74.4	48.8	21.3	61.0	27.8	15.9	2.4	27.2	49.6	51.2
3	74.6	46.9	38.7	60.1	29.1	16.2	2.5	27.5	49.3	51.5
4	74.6	45.0	27.5	60.9	30.5	16.4	2.8	27.2	49.8	50.8
5	74.2	48.1	35.6	59.5	29.0	15.6	2.5	26.8	49.6	51.2
6	74.7	48.8	26.9	60.0	29.1	16.2	3.2	27.8	50.1	51.8
7	74.5	43.8	33.1	58.4	27.4	16.0	2.7	27.7	49.6	51.4
8	74.4	51.9	33.8	62.7	28.4	16.6	2.7	27.1	49.5	51.4
9	74.3	51.2	30.6	61.9	30.9	16.2	2.8	28.0	49.7	51.1
10	74.5	50.6	27.5	60.1	28.3	15.5	2.8	27.1	49.8	50.9

C.2.3 DeepAug ERM (CAE) [15]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	73.9	52.5	38.1	62.0	31.1	13.3	2.6	29.3	63.5	57.1
2	73.7	55.6	38.7	63.7	31.4	13.1	2.4	28.0	63.0	56.0
3	73.3	50.6	33.8	64.5	29.5	13.7	2.5	27.6	63.2	56.4
4	73.7	48.8	30.6	65.9	30.3	13.1	2.5	28.9	63.2	56.2
5	73.6	50.6	27.5	63.7	30.1	13.3	2.7	27.8	63.2	56.9
6	73.8	53.8	39.4	66.5	31.4	13.2	2.8	28.8	63.3	57.1
7	73.6	52.5	42.5	65.6	31.2	13.7	2.9	28.4	63.3	57.3
8	73.6	55.0	38.1	62.6	30.8	13.1	2.5	28.0	63.2	57.1
9	73.5	48.1	40.0	62.6	29.6	13.3	2.6	28.1	62.9	56.2

C.2.4 DeepAug ERM (EDSR) [15]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	72.7	51.2	34.4	63.5	31.8	10.9	2.0	26.5	52.3	65.0
2	72.7	52.5	30.0	59.8	32.9	10.6	2.0	26.4	52.3	65.1
3	72.9	53.8	28.7	60.1	29.5	11.3	1.9	26.8	52.6	65.3
4	73.0	50.0	28.7	61.3	32.0	11.0	1.9	26.2	52.6	65.2
5	73.2	50.6	39.4	62.6	35.2	10.8	2.3	27.5	52.6	65.3
6	73.2	51.2	23.7	62.7	31.6	11.3	1.9	26.6	52.5	65.1
7	72.6	51.2	31.2	59.9	33.1	10.7	2.2	26.8	52.3	64.8
8	72.7	53.1	30.0	58.6	31.5	11.0	1.8	25.5	51.8	65.2
9	72.6	51.9	33.1	61.1	32.8	11.6	1.9	26.1	52.3	65.0
10	72.9	50.0	37.5	61.5	33.4	11.0	2.4	26.7	52.0	65.0

C.2.5 Stylized ImageNet [13]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	56.7	47.5	63.1	71.2	53.4	53.2	0.9	25.5	40.1	40.9
2	56.1	47.5	59.4	70.2	55.2	53.5	0.8	24.9	39.0	40.4
3	55.3	41.2	60.0	69.1	55.3	53.1	0.7	24.9	39.0	40.1
4	55.9	45.6	58.1	69.0	52.5	53.3	0.9	25.0	39.0	40.3
5	56.1	45.0	56.2	72.0	51.4	53.6	0.8	25.4	38.8	40.5
6	55.7	50.6	61.3	70.1	54.5	53.1	0.7	24.8	39.0	39.9
7	55.7	45.0	58.1	71.6	53.8	53.1	0.8	24.8	38.4	40.0
8	56.3	47.5	55.6	71.0	54.1	53.3	0.8	25.1	39.5	40.9
9	56.0	48.8	60.0	67.5	54.9	53.4	0.6	25.1	39.2	40.5
10	55.2	50.0	52.5	69.7	51.6	52.8	0.7	24.5	38.8	40.0

C.2.6 InfoDrop [37]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	73.7	45.0	25.0	56.5	23.3	7.8	2.5	23.2	44.4	48.4
2	73.1	49.4	20.0	58.5	22.4	7.9	2.2	22.8	43.9	48.7
3	73.2	48.1	21.3	57.6	22.4	7.6	2.2	22.4	44.4	48.5
4	73.4	50.0	16.2	57.4	22.6	7.6	2.3	22.6	44.6	49.5
5	73.1	50.0	20.0	55.8	22.7	8.0	2.2	23.0	44.8	48.6
6	73.7	45.6	16.2	60.5	21.8	7.6	2.3	23.0	44.5	49.3
7	73.3	49.4	12.5	54.3	23.8	7.8	2.0	22.2	44.4	48.7
8	73.6	47.5	23.7	56.0	23.4	7.9	2.2	22.5	44.3	48.2
9	73.3	50.6	13.1	53.9	23.2	7.8	2.2	22.2	44.3	48.7
10	73.1	41.9	21.9	56.6	22.5	8.6	2.2	22.7	44.4	47.7

C.2.7 SagNet [31]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	75.2	43.8	29.4	59.2	19.8	6.1	1.6	22.3	44.4	48.2
2	73.9	44.4	24.4	58.4	19.8	6.4	1.8	22.1	43.2	47.2
3	73.9	41.9	28.1	59.6	20.4	6.1	1.3	21.4	43.5	47.3
4	73.9	45.0	23.1	59.2	19.8	6.2	1.4	21.7	44.0	47.6
5	73.9	41.9	23.7	58.7	19.7	6.2	1.5	21.7	43.8	47.3
6	75.1	45.6	25.0	58.0	20.3	6.3	1.8	22.2	44.2	48.8
7	73.9	45.0	26.2	62.0	19.7	6.2	1.9	22.5	43.3	47.5
8	74.0	42.5	24.4	59.2	20.0	6.0	1.2	21.3	44.1	47.8
9	73.9	41.2	24.4	59.5	19.9	6.1	1.4	21.7	43.9	47.9
10	74.0	41.9	24.4	58.2	20.2	6.3	1.5	21.1	43.8	47.5

C.2.8 pAdaIN [34]

Run	ImageNet1k	Silhouette	Edge	Sketch	CueConflict	ImageNetStylized	ImageNetA	ImageNetR	DeepAugCAE	DeepAugEDSR
1	73.0	44.4	21.3	56.0	21.3	8.3	1.4	21.3	43.0	49.3
2	73.2	48.1	13.1	56.7	21.6	8.1	1.5	21.7	42.8	48.6
3	73.1	41.2	23.7	56.2	20.5	8.1	1.3	21.1	42.7	48.6
4	73.0	44.4	21.3	56.0	21.3	8.3	1.4	21.3	43.0	49.3
5	73.2	42.5	18.8	57.9	20.9	8.1	1.7	22.0	42.5	48.3
6	73.1	46.3	24.4	57.1	22.3	8.2	1.4	21.3	43.2	48.7
7	73.3	43.1	22.5	55.8	21.3	8.1	1.6	22.1	42.6	48.6
8	73.3	45.6	20.0	58.0	21.2	8.0	1.4	21.2	42.9	48.5
9	73.2	45.6	24.4	56.9	22.0	8.2	1.6	21.6	42.9	48.9
10	73.3	40.0	21.3	57.1	21.7	8.0	1.4	21.2	42.5	48.6