6 Supplementary Material

6.1 Notations

We conclude all the notations we used in the paper in Table 3. Specifically, the notations with superscription "s" and "t" indicate they are from source domain and target domain, respectively.

Notation	Description
α	Neural architecture parameters
w	Neural network weights
Φ	Machine learning model
P/Q	Source / Target distribution
G	Feature generator in phase II
C	Classifier in phase II
$\mathcal{L}_{train}$ / $\mathcal{L}_{val}$	Training / Validation loss
$\mathbf{x}$ ($\mathbf{x}^s$, $\mathbf{x}^t$)	Training data (source, target)
$\mathbf{y}$ ($\mathbf{y}^s$, $\mathbf{y}^t$)	Labels (source, target)
$\mathcal{D}_s$ / $\mathcal{D}_t$	Source / Target Domain
$o^{(i,j)}$ / $e^{(i,j)}$	Operation / Edge from node i to j in NAS Graph
$x^{(i)}$	i-th node in a neural cell
L	Number of nodes in a neural cell
d_k^2 / $\hat{d}_k^2$	MK-MMD / Empirical MK-MMD
k / $\mathcal{K}$	Kernel / Kernels
m	Number of data used to compute MK-MMD
ξ	Learning rate of inner optimization
λ	Trade-off parameters
K	Class number
$p(y\|x)$	Probabilistic outputs of classifiers
N	Number of Classifiers

Table 3: We conclude all the notations we used in the paper.

6.2 Data Statistics

The datasets used in this paper are described in Table 4. Specifically, the **USPS**, **MNIST**, **SVHN**, **CIFARO-10**, **STL** datasets are from TORCHVISION[1]. The **SYN SIGNS**[2] and GTSRB [3] dataset are downloaded from their official websites.

	# train	# test	# classes	Target	Resolution	Channels
USPS	7,291	2,007	10	Digits	16×16	Mono
MNIST	60,000	10,000	10	Digits	28×28	Mono
SVHN	73,257	26,032	10	Digits	32×32	RGB
CIFAR-10	50,000	10,000	10	Object ID	32×32	RGB
STL	5,000	8,000	10	Object ID	96×96	RGB
SYN SIGNS	100,000	–	43	Traffic signs	40×40	RGB
GTSRB	32,209	12,630	43	Traffic signs	*varies*	RGB

Table 4: Statistics of datasets we used in our paper.

Data Preparation Some of the experiments that involved datasets described in Table 4 required additional data preparation in order to match the resolution and format of the input samples and match the classification target. These additional steps will now be described.

STL $\rightarrow$ CIFAR-10 CIFAR-10 and STL are both 10-class image datasets. The STL images were down-scaled to 32×32 resolution to match that of CIFAR-10. The 'frog' class in CIFAR-10 and the

[1]https://pytorch.org/docs/stable/torchvision/datasets.html

[2]http://graphics.cs.msu.ru/en/node/1337

[3]http://benchmark.ini.rub.de/?section=gtsrb

'monkey' class in STL were removed as they have no equivalent in the other dataset, resulting in a 9-class problem with 10% less samples in each dataset.

Syn-Signs $\rightarrow$ GTSRB GTSRB is composed of images that vary in size and come with annotations that provide region of interest (bounding box around the sign) and ground truth classification. We extracted the region of interest from each image and scaled them to a resolution of 40×40 to match those of Syn-Signs.

SVHN $\rightarrow$ MNIST The MNIST images were padded to 32×32 resolution and converted to RGB by replicating the greyscale channel into the three RGB channels to match the format of SVHN.

6.3 ML REPRODUCIBILITY

We have submitted the code for all the experiments in the supplementary material. We will briefly describe the details about the experiments.

Hyper-parameter We set λ to be 1 for all the experiments. The range of learning rate we considered is between 2e-4 to 0.25. We adopt grid search to select the best hyper-parameters. All the hyper-parameters used to generate results can be viewed in the code.

Measure For all the quantitative results in the paper, we use accuracy as the measurement.

Average runtime For Phase I in our model, *i.e.* searching the neural architecture, our model takes 0.3 GPU day to find the optimal architecture. For Phase II, our model typically takes about two days to converge.

Computing infrastructure Our code is based on Pytorch 1.2.0 and Torchvision 0.4.2. All other descriptions can be found in the readme file in the code.