Supplementary Material: How Far Pre-trained Models Are from Neural Collapse on the Target Dataset Informs their Transferability

Zijian Wang[1] Yadan Luo[1] Liang Zheng[2] Zi Huang[1] Mahsa Baktashmotlagh[1]
[1]The University of Queensland [2]Australian National University
{zijian.wang, y.luo, helen.huang, m.baktashmotlagh}@uq.edu.au, liang.zheng@anu.edu.au

Reproducibility: The code of this project will be available at https://github.com/BUserName/NCTI

1. Dataset Description

FGVC Aircraft is a fine-grained aircraft classification dataset with 10,000 images that belong to 100 categories. The training split of the dataset contains 6,667 images, and the test split contains 3,333 images.

Stanford Cars is a fine-grained car classification dataset with 16,815 images in 196 classes. The training set consists of 8,144 images, and the test set consists of 8,041 images.

Food-101 is a fine-grained food classification dataset that comprises 101,000 pictures that are classified into 101 categories of food. Within each food category, there are 750 images for training purposes and 250 images for testing purposes.

Oxford-IIIT Pets is a fine-grained pet dataset that contains 7,049 pet images in the dataset, belonging to 47 different species. The training set includes 3,680 images while the testing set has 3,669 images.

Oxford-102 Flowers is a fine-grained flower classification dataset that consists of 102 categories of flowers. Each category contains between 40 and 258 images. We sample 20 images per category to construct the training set and use the rest of the 6,149 images as the testing set.

Caltech101 is a coarse-grained classification dataset with 9,146 images that belong to 101 categories. We sample 70% of data as the training set.

CIFAR-10 and **CIFAR-100** are coarse-grained classification datasets that consist of 60,000 color images, each measuring 32x32 pixels, and divided into 10 distinct classes. For each class, CIFAR-10 has 5,000 images allocated for training and 1,000 images for testing, while CIFAR-100 has 500 images allocated for training and 100 images for testing

VOC2007 is a coarse-grained classification dataset that has 9,963 images in 20 classes. Following the official split, we have 5,011 training images.

SUN397 is a scene classification dataset that has 397 classes, each having 1,000 scenery pictures. Unless otherwise specified, we use the official training split 1 (19,850 images) to perform the estimation.
DTD is a texture classification dataset that includes 5,640 textural images, ranging in size from 300x300 to 600x600. The dataset is categorized into 47 classes, and each class contains 80 training images and 40 testing images.

2. More Correlation Results

We compare the correlation between scores of different transferability metrics and model recognition performance (%) in Fig. 1 and 2. We can see that our method demonstrates strong correlations between the calculated score and the recognition performance.

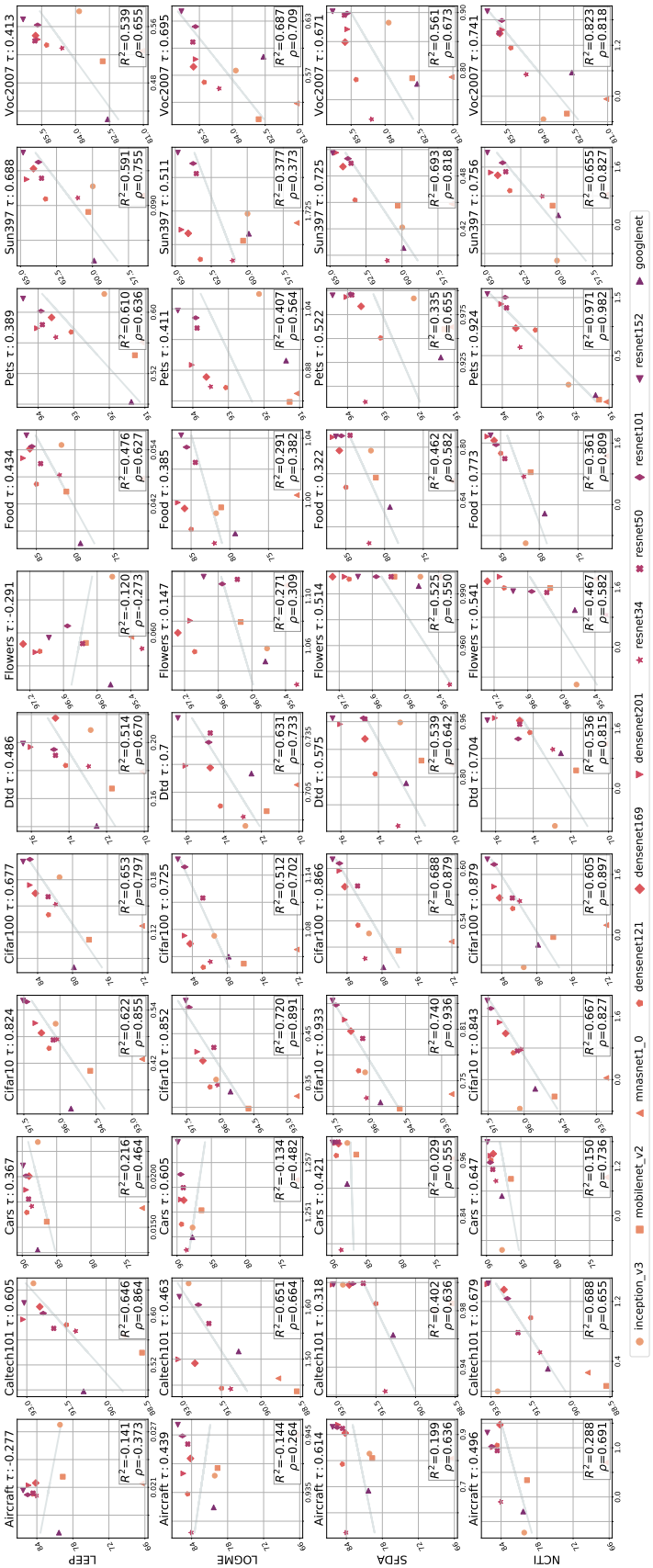

Figure 1. Correlation between transferability scores and model recognition performance (%) on the target dataset after fine-tuning. Each marker represents a different supervised pre-trained model. We show R-square value R^2, Spearman's coefficient ρ, and weighted Kendall's τ here.

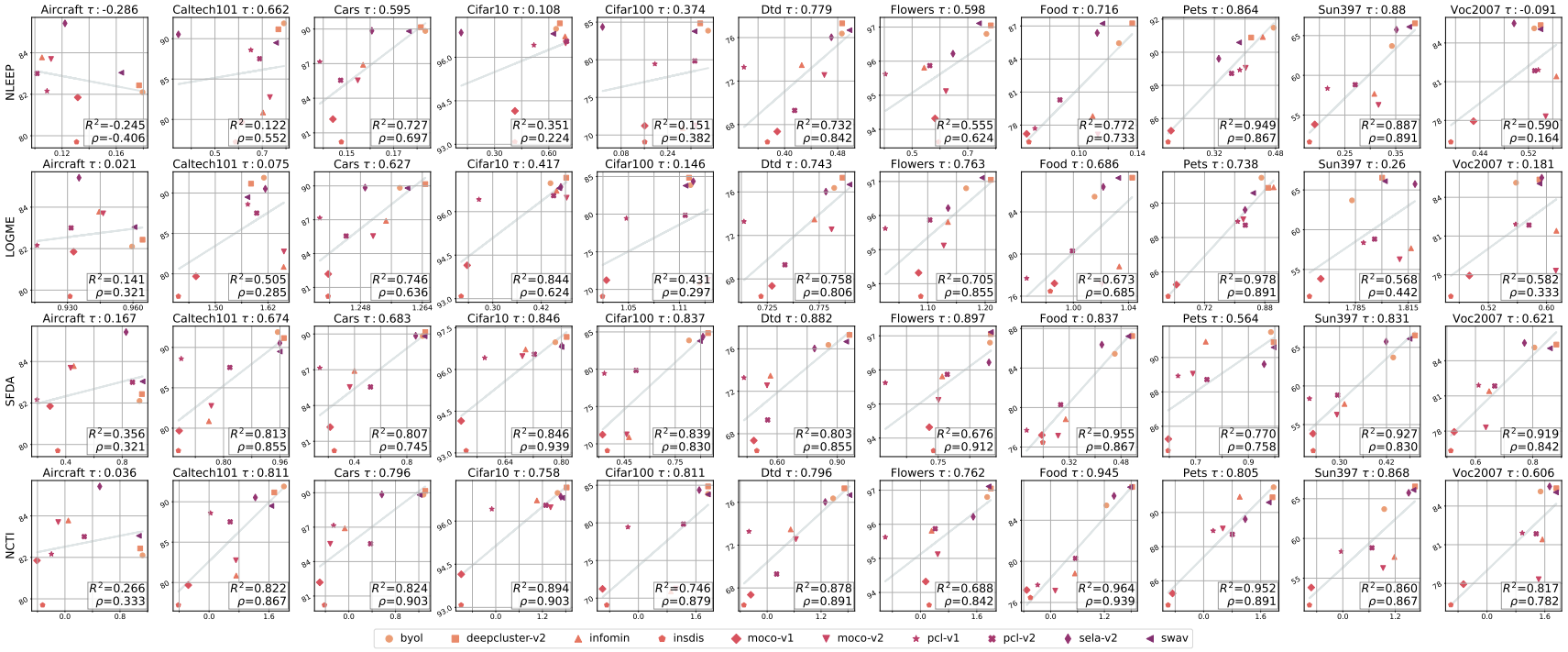

Figure 2. Correlation between transferability scores and model recognition performance (%) on the target dataset after fine-tuning. Each marker represents a different self-supervised pre-trained model. We show R-square value R^2, Spearman's coefficient ρ, and weighted Kendall's τ here.