Scaling Laws for the Out-of-Distribution Generalization of Image Classifiers

Gabriele Prato [1 2] **Simon Guiroy** [* 1 2] **Ethan Caballero** [* 1 2] **Irina Rish** [1 2] **Sarath Chandar** [1 3]

Abstract

We empirically study the effects of scale on the out-of-distribution generalization of widely used image classification models. Three parameters are independently scaled: training set size, number of training classes and model size. On 70 train-target pairs, we find that in general scaling the amount of training data improves out-of-distribution (OOD) performance, while scaling the amount of training classes helps, but up to a point. When scaling model size, not all architectures improve OOD, with some getting mixed results. We also find that increasing the size of natural image datasets seems to improve OOD generalization on natural datasets almost universally. For example, training on more fungi data results in models better at classifying aircrafts. Additionally, scaling natural image datasets improves OOD performance on nonnatural image datasets, but the same cannot be said about the inverse. When scaling the number of classes, we find that increasing the count of diverse classes seems more beneficial than increasing classes in a specific domain. Finally, we find power laws for a great number of train-target pairs and architectures. Our findings shed new light on the relationship between scaling and OOD generalization.

1. Introduction

Over the past few years, deep learning has made tremendous progress in fields such as vision (Alam et al., 2020) and natural language processing (Torfi et al., 2020). While these advances are impressive, neural networks still have many issues, the biggest possibly being that they are known to perform poorly out-of-distribution (Arjovsky, 2021). One particular problem models can encounter is simultaneous changes in the data distribution $p(x)$ and task $p(y|x)$ at test time. For example, a self-driving car seeing an elephant for the first time should be able to recognize it as something new (out-of-distribution detection) and then seeing another elephant, it should be able to recognize it as being the same thing as the one previously seen (out-of-distribution generalization). If neural networks are to be successfully deployed in the real world, then they most likely need to be able to deal well with situations not seen during training. One possible way to improve OOD generalization is scale. Indeed some very large networks do seem to manifest a certain ability to generalize out-of-distribution (Brown et al., 2020; Radford et al., 2021; Ramesh et al., 2021), but more analysis in this direction needs to be performed.

The Deep Learning book (Goodfellow et al., 2016) states that: "for nonparametric models, more data yield better generalization until the best possible error is achieved." While neural networks are usually parametric models, practitioners tend to adjust their capacity for a given problem based on performance. This is a form of nonparametric learning algorithm. The statement holds because of the i.i.d. assumption, but what happens when samples from the train and test set are not identically distributed? Can more training samples still help when test samples are out-of-distribution? According to the no free lunch theorem (Wolpert & Macready, 1997), not if we consider all possible data-generating distributions. In practice though, assumptions can be made about the kind of probability distributions that we encounter, since we are interested in real world problems. So what happens in such cases, does more training data help?

One real world domain where assumptions can be made about the probability distribution is that of vision. Scaling on real world vision tasks can potentially result in out-of-distribution generalization on other real world vision problems. Because of its widespread use, vision is a very interesting case study. It's a simple, yet very important problem and a very good example of real world application. Indeed, it is hard to imagine neural networks being deployed in the real world without the need of vision. Agents which do use vision will quite certainly encounter situations not seen during training. Knowing how these models will perform in such cases is very important.

In this work, we empirically study the effect of scaling various aspects of the training data as well as model size

[*]Equal contribution [1]Mila [2]Université de Montréal [3]École Polytechnique de Montréal. Correspondence to: Gabriele Prato <prato-gab@mila.quebec>.

Presented at the ICML 2021 Workshop on Uncertainty and Robustness in Deep Learning.,

on the generalization error. Specifically, we are interested in the effect of scale in the vision domain. For our study, we train standard image classifiers on various datasets for multiple different values of our scaling parameters. We then measure their out-of-distribution performance using multiple datasets. Models are never fine-tuned on the target datasets. At test time, the classification layer is discarded and the models are evaluated as prototypical networks (Snell et al., 2017), all weights remaining frozen.

Our empirical study is extensive. While typically only model size and amount of training data are scaled, we also analyze the impact of scaling the number of training classes. All three scaling parameters are studied independently, meaning that when one is scaled, the others remain fixed. Our experiments are performed on widely used architectures in the field of vision, trained and tested on ten datasets, ranging from natural images to drawings. In total, our study encompasses 70 diverse train-target pairs, which amounts to thousands of experiments.

By performing this study, our goal is to shed some light on the question of how much scale can improve out-of-distribution generalization and provide useful insight for future work directions.

2. Experimental Setup

In this section, we detail the training and evaluation procedure common to all experiments as well as the power laws we fit.

2.1. Training & Evaluation Procedure

In all cases, models are trained as standard image classifiers and evaluated as prototypical networks. Best OOD performance is reported for each model, measured throughout training. See the following section for a detailed explanation of prototypical networks.

The ten datasets from Meta-Dataset (Triantafillou et al., 2019) are used in our study: Aircraft, Bird, COCO, Describable Texture, Flower, Fungi, ImageNet, Omniglot, Quickdraw and Traffic Sign. 80% of each dataset's training classes are used as the training set, the other 20% is kept for OOD evaluation when the target dataset is the same as the train dataset, e.g. Aircraft-Aircraft. When the target dataset is different than the train dataset, we evaluate on all classes and all data of that target dataset.

We perform 5-way 1-shot and 5-way 5-shot OOD evaluation, repeated over thousands of trials. For each trial, a set of five classes is chosen at random and a prototype for each class in the set is computed with one sample in the case of 1-shot and five samples in the case of 5-shot. Then, samples are drawn at random from each class in the set and classified using the prototypical network method. We found 5-way 1-shot results to be slightly more noisy than 5-way 5-shot.

2.2. Prototypical Networks

Instead of outputting a logit for each class, prototypical networks (Snell et al., 2017) compute a metric between the model's output and each class prototype. The result is then used to predict to which class the input sample belongs to. Typically, as is our case, the classification layer is discarded and the final latent representation is used as the output instead. Prototypes are computed by averaging the output embedding of same class samples from a support set. Common metric choices are euclidean distance and cosine similarity, although any function could be used. We use euclidean distance in our study.

The advantage of prototypical networks is that they are not constrained to making predictions for samples only belonging to the training classes. Prototypes can be computed with samples from classes never seen during training. This allows them to perform classification on any set of classes, making it very practical to evaluate a model's classification performance out-of-distribution.

2.3. Power laws

For each train-target pair and each model, we fit when possible simple power laws for both the scaling of the training set size N and the number of training classes C:

$$Acc(X) = c + x_0 X^{\alpha} \tag{1}$$

where X can be N or C, Acc is the estimated accuracy and c, x_0 and α are constants found via the least squares method. Power laws where omitted from Figures in the following section to ease reading. All power laws are reported with larger plots in Appendix section B.

3. Empirical Results

The following sections detail our findings for each scaled parameter. Full results are available in the Appendix section B.

3.1. Scaling Amount of Training Data

The first parameter that we scale is the amount of training data. Five ratios are used: 100%, 50%, 25%, 12.5% and 6.25% of the total training data for each dataset. The amount of training classes remains the same no matter the ratio. Because of Flower's dataset size, we only trained with 100%, 50%, 25% and 12.5% of the data. For each ratio, we train five architectures: a four-layer convolutional network (LeCun et al., 1989), a VGG11 (Simonyan & Zisserman, 2014), a ResNet18 (He et al., 2015), a DenseNet121 (Huang

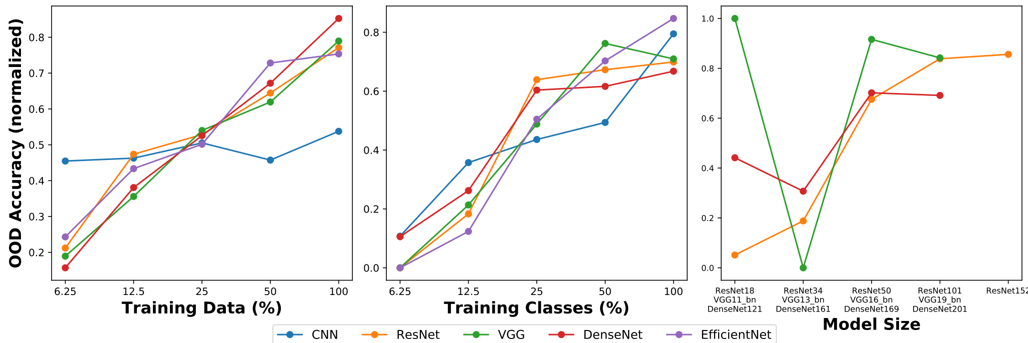

Figure 1. OOD accuracy averaged across all 70 train-target pairs in our study when scaling three parameters: amount of training data (left), number of training classes (center) and model size (right). The first two x-axis are a percentage of the maximum value (per dataset) and the third lists the architecture variants. Y-axis is the 5-way 5-shot normalized accuracy. See section B for 5-way 1-shot results. Performance is normalized to show trend. For unnormalized performance, including power laws, see Figure 5.

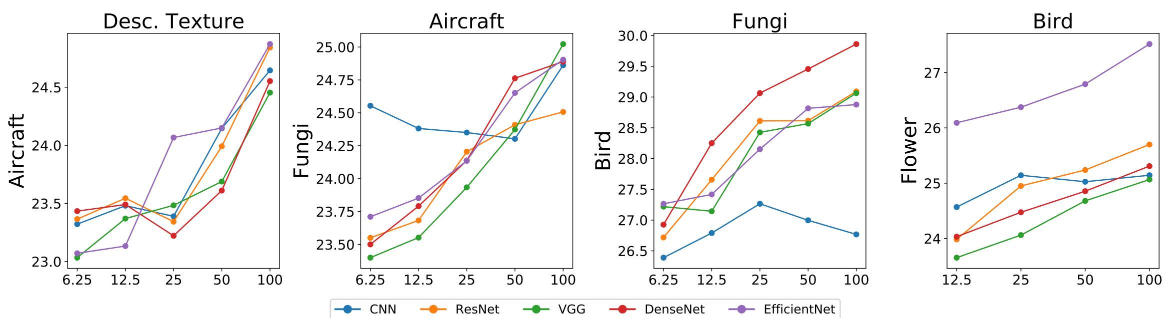

Figure 2. OOD Accuracy for multiple train-target pairs of natural image datasets. For each plot, the train dataset is written on the left and the target dataset on top. X-axis is the percentage of the total training data and y-axis is the 5-way 5-shot accuracy. For 5-way 1-shot accuracy, see section B.

et al., 2016) and an EfficientNet B0 (Tan & Le, 2019). All five models are trained on all datasets and evaluation is performed on Aircraft, Bird, Describable Texture, Flower, Fungi, Omniglot and Traffic Sign. Evaluation on COCO, ImageNet and Quickdraw will be done in the future, as well as OOD evaluation on the same dataset as the training dataset.

On average, see Figure 1, when scaling the amount of training data, we find that the out-of-distribution performance of the image classifiers improves. The CNN is the only model for which the 5-way 1-shot performance does not seem to improve on average (see section B). Even for 5-way 5-shot, the CNN's rate of improvement is not on par as the other models which all follow a very similar trend. We believe this might be due to the CNN being more shallow than other networks and it's classification layer playing a more important role due to it being very large. We hypothesize that by removing it when performing the prototypical network evaluation, performance is severely handicapped.

On average, when the training dataset is one of natural images (Aircraft, Bird, COCO, Describable Texture, Flower, Fungi and ImageNet), the out-of-distribution generalization on other natural image datasets does seem to improve, see Figure 2. We find this very interesting, as it is not so obvious if training on more fungi for example can make a model better at classifying aircrafts. If it turns out that training on more natural images almost universally improves the out-of-distribution generalization on any natural image dataset, then scale could potentially be a good solution or at least part of the solution to the OOD problems that models deployed in the real world face.

As for out-of-distribution performance on nonnatural image datasets, surprinsingly, scaling the training datasets such as Aircraft, Bird, Describable Texture, Flower and Fungi results in better performance on Omniglot, a character dataset, see Figure 6. This shows that training on bigger natural image datasets can also potentially improve the out-of-distribution generalization on nonnatural image datasets.

This also highlights the unintuitivity in predicting whether or not training on more data from a certain domain will result in better OOD generalization on some other domain. Indeed, we expected that for the OOD generalization to improve, distributions would have to be much closer. Inversely though, training on more Omniglot data does not seem on average to provide as much gains on other datasets, see Figure 6. This could mean that while natural image datasets provide good generalization to all sorts of domains aside from natural images, the inverse might not provide the same benefits, at least not to the same degree. More experiments would have to be conducted between natural and nonnatural image datasets to verify this hypothesis.

The only training dataset where on average scaling does not seem to improve the out-of-distribution generalization is the Traffic dataset. As an evaluation dataset, apart from a few exceptions, OOD performance on Traffic does not seem to improve much as training data is scaled. We believe these issues might be due to the nature of the task, where many classes share key features as well as potentially the low quality of the data. See section B for results.

3.2. Scaling Amount of Training Classes

In the last section, the number of training classes was fixed for all training data ratios. We now take a look at the inverse setup, where the amount of training classes scales, while the amount of training data remains the same. In this setup, we are interested in seeing if more classes helps with generalization. We use four ratios for our experiments: 100%, 50%, 25% and 12.5%. 6.25% is left out due to the low number of classes in most datasets. Again, we train the same five architectures on all ten datasets for all ratios and evaluate on the same seven datasets as in section 3.2.

On average, see Figure 1 and 5, when scaling the amount of training classes, OOD performance seems to improve at lower percentages, but then plateau or start to plateau at some point, the only exception being the CNN. Estimated power laws in Figure 5 for the ResNet and DenseNet models would potentially be better with more data points to fit the function.

Omniglot, Fungi and ImageNet are the datasets with the most number of train classes, 1296, 1120 and 800 respectively. Scaling the amount of training classes for Fungi and Omniglot does not seem on average to improve the OOD generalization. For ImageNet though, more classes does seem to help, but the gains seem to level off after more than 200 classes. See Figure 7.

COCO and ImageNet are the datasets with the most diverse set of classes. While gains when scaling on ImageNet seem to flatten towards the end, for COCO scaling seems to improve the OOD generalization on every target dataset, see Figure 7. COCO only has a total of 64 train classes though, which could explain why the performance keeps improving. When the number of classes is low in the ImageNet experiments, we also see steady gains on average.

For other datasets, we find results to be mixed. In general though, it seems like when the number of classes is low, more is better, but up to a certain point. It also seems like more from a diverse set of classes helps more than more classes from a more specific domain.

3.3. Scaling Model Size

Finally we scale the model size. Changing model capacity can lead to underfitting or overfitting and hurt generalization. For this experiment though, we are interested in analyzing scaling of neural networks that practitioners typically use on large datasets and how their OOD performance varies.

The architectures that we train are the following: ResNet18, ResNet34, ResNet50, ResNet101, ResNet152, a VGG11, VGG13, VGG16, VGG19 and a DenseNet121, DenseNet161, DenseNet169, DenseNet201. EfficientNet scaling will be done in the future. It is well known that when scaling these models, their in-distribution performance improves for standard large image classification datasets such as ImageNet (Simonyan & Zisserman, 2014; He et al., 2015; Huang et al., 2016). The question is if the OOD generalization follows the same trend. For our experiment, we train the models on ImageNet and evaluate on Aircraft, Bird, Describable Texture, Flower, Fungi, Omniglot and Traffic Sign. Apart from the model size, no other parameter is scaled.

On average, see Figure 1, we see that only the ResNet model improves each time it is scaled. For DenseNet, the last two variants performed better on average then the first two, but to a lesser degree than the ResNet. Furthermore, looking at the 5-way 1-shot results for the DenseNet, section B, its OOD generalization does not improve as it scales. For VGG, in neither the 5-way 1-shot nor 5-way 5-shot experiments did the OOD generalization improve, see section B.

Looking at individual train-target pairs, Figure 9, we see that the trend is similar in all cases. The VGG performance varies by a lot, while the ResNet improves in most cases as it scales and as for the DenseNet, there are no clear trends.

4. Conclusion

In this work, we studied the effect of scale on the out-of-distribution generalization of image classifiers. We found that in general, scaling the training set size improves OOD performance. We also found that scaling the amount of training classes seems to help when the number of classes is low, but saw diminishing returns for higher values. As for scaling model size, we found the results to be mixed.

References

Alam, M., Samad, M. D., Vidyaratne, L., Glandon, A., and Iftekharuddin, K. M. Survey on deep neural networks in speech and vision systems. *Neurocomputing*, 417: 302–321, 2020.

Arjovsky, M. Out of Distribution Generalization in Machine Learning. *arXiv e-prints*, art. arXiv:2103.02667, March 2021.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. Language Models are Few-Shot Learners. *arXiv e-prints*, art. arXiv:2005.14165, May 2020.

Djolonga, J., Yung, J., Tschannen, M., Romijnders, R., Beyer, L., Kolesnikov, A., Puigcerver, J., Minderer, M., D'Amour, A., Moldovan, D., Gelly, S., Houlsby, N., Zhai, X., and Lucic, M. On Robustness and Transferability of Convolutional Neural Networks. *arXiv e-prints*, art. arXiv:2007.08558, July 2020.

Goodfellow, I., Bengio, Y., and Courville, A. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative Adversarial Networks. *arXiv e-prints*, art. arXiv:1406.2661, June 2014.

He, K., Zhang, X., Ren, S., and Sun, J. Deep Residual Learning for Image Recognition. *arXiv e-prints*, art. arXiv:1512.03385, December 2015.

Henighan, T., Kaplan, J., Katz, M., Chen, M., Hesse, C., Jackson, J., Jun, H., Brown, T. B., Dhariwal, P., Gray, S., Hallacy, C., Mann, B., Radford, A., Ramesh, A., Ryder, N., Ziegler, D. M., Schulman, J., Amodei, D., and McCandlish, S. Scaling Laws for Autoregressive Generative Modeling. *arXiv e-prints*, art. arXiv:2010.14701, October 2020.

Hernandez, D., Kaplan, J., Henighan, T., and McCandlish, S. Scaling Laws for Transfer. *arXiv e-prints*, art. arXiv:2102.01293, February 2021.

Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., Patwary, M. M. A., Yang, Y., and Zhou, Y. Deep Learning Scaling is Predictable, Empirically. *arXiv e-prints*, art. arXiv:1712.00409, December 2017.

Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. Densely Connected Convolutional Networks. *arXiv e-prints*, art. arXiv:1608.06993, August 2016.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., and Amodei, D. Scaling Laws for Neural Language Models. *arXiv e-prints*, art. arXiv:2001.08361, January 2020.

Kingma, D. P. and Welling, M. Auto-Encoding Variational Bayes. *arXiv e-prints*, art. arXiv:1312.6114, December 2013.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.*, 1(4):541–551, December 1989. ISSN 0899-7667. doi: 10.1162/neco.1989.1.4.541. URL http://dx.doi.org/10.1162/neco.1989.1.4.541.

Lu, K., Grover, A., Abbeel, P., and Mordatch, I. Pretrained Transformers as Universal Computation Engines. *arXiv e-prints*, art. arXiv:2103.05247, March 2021.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I. Learning Transferable Visual Models From Natural Language Supervision. *arXiv e-prints*, art. arXiv:2103.00020, February 2021.

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., and Sutskever, I. Zero-Shot Text-to-Image Generation. *arXiv e-prints*, art. arXiv:2102.12092, February 2021.

Simonyan, K. and Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv e-prints*, art. arXiv:1409.1556, September 2014.

Snell, J., Swersky, K., and Zemel, R. S. Prototypical Networks for Few-shot Learning. *arXiv e-prints*, art. arXiv:1703.05175, March 2017.

Tan, M. and Le, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv e-prints*, art. arXiv:1905.11946, May 2019.

Torfi, A., Shirvani, R. A., Keneshloo, Y., Tavaf, N., and Fox, E. A. Natural Language Processing Advancements By Deep Learning: A Survey. *arXiv e-prints*, art. arXiv:2003.01200, March 2020.

Triantafillou, E., Zhu, T., Dumoulin, V., Lamblin, P., Evci, U., Xu, K., Goroshin, R., Gelada, C., Swersky, K., Manzagol, P.-A., and Larochelle, H. Meta-Dataset: A Dataset of Datasets for Learning to Learn from Few Examples. *arXiv e-prints*, art. arXiv:1903.03096, March 2019.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention Is All You Need. *arXiv e-prints*, art. arXiv:1706.03762, June 2017.

Wolpert, D. H. and Macready, W. G. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 1997.

Zhao, S., Ren, H., Yuan, A., Song, J., Goodman, N., and Ermon, S. Bias and Generalization in Deep Generative Models: An Empirical Study. *arXiv e-prints*, art. arXiv:1811.03259, November 2018.

A. Related Work

Recently there has been a gain in interest in studying the effects of scale in machine learning, although only few studies have been done so far. Hestness et al. (2017) are the first to study the manner in which deep learning scales predictably empirically via a power law functional form. They specifically study how IID validation loss scales with respect to training dataset size on language modelling, image classification, and speech recognition tasks.

Zhao et al. (2018) perform multiple experiments where they isolate features such as number of element and color and check if GANs (Goodfellow et al., 2014) and VAEs (Kingma & Welling, 2013) manage to generalize or not. For example, when training with images containing always three objects, they show that generative models do generate samples containing fewer, equal and more objects. They show the same thing happening for color and when mixing these features: "We find that if the training distribution only contains a small number of combinations (e.g., 10-20) (in a feature space Z) the learned distribution memorizes them almost exactly. However as there are more combinations in the training set, the model starts to generate novel ones."

Kaplan et al. (2020) study scaling laws for Transformer (Vaswani et al., 2017) language models. They find that the in-distribution performance has a power law relationship with the dataset size, model size and amount of training individually when not bottlenecked by the other two parameters. Additionally, for an equal number of parameters, tuning within a reasonable range the Transformer's depth versus width has little impact on performance. They also demonstrate that overfitting follows a simple ratio between model and dataset size. They show that larger models are more sample efficient and that when training compute is on a budget, but model and dataset size is not, optimal performance is obtained by training very large models and early stopping, rather than training smaller models to convergence. Finally, when out-of-distribution, they find that the language modelling performance correlates with the validation performance of the training problem, offset by a constant. Brown et al. (2020) train language models of various sizes, then evaluate them on multiple downstream natural language processing tasks without any fine-tuning and restrictions on the amount of training data or training compute budget. They show that performance scales in relationship with model size.

Djolonga et al. (2020) evaluate the impact of model size and dataset size on robustness, where classes at train and test time are the same, but there is a distribution shift in the data, for example changes in the lighting of image samples. They find that scaling both model size and training set size improves such robustness. Henighan et al. (2020) train autoregressive Transformers on tasks from different domains, including vision, language and math. The latter is the only domain for which they evaluated outside of the training distribution. Models were evaluated on increasingly difficult problems beyond the difficulty of the training set. In all cases they find that loss scales with respect to model size and amount of training.

Radford et al. (2021) train a model to match images with their captions. They vary the amount of compute and show that zero-shot error averaged over multiple vision datasets scales as a function of compute. Hernandez et al. (2021) use language models pretrained on text and fine-tuned on python code to show that when the fine-tuning dataset is small, models trained from scratch eventually hit a performance wall no matter the model size, while equivalently sized pre-trained models keep improving. They also show that the smaller the fine-tune dataset, the more larger models help when pre-trained compared to training from scratch. To reach a certain loss, pre-trained models need far less training examples than models trained from scratch. This difference scales with respect to model size and inversely with the size of the fine-tuning dataset.

Lu et al. (2021) pretrain a Transformer with a language modelling task. Then they freeze the feed-forward and self-attention weight and finetune the rest of the model on non-language tasks such as image classification, protein folding and numerical computation. They show that even with such a drastic change in modality, the models with frozen weights perform just as well as Transformers trained from scratch on each individual task.

B. Full Empirical Results

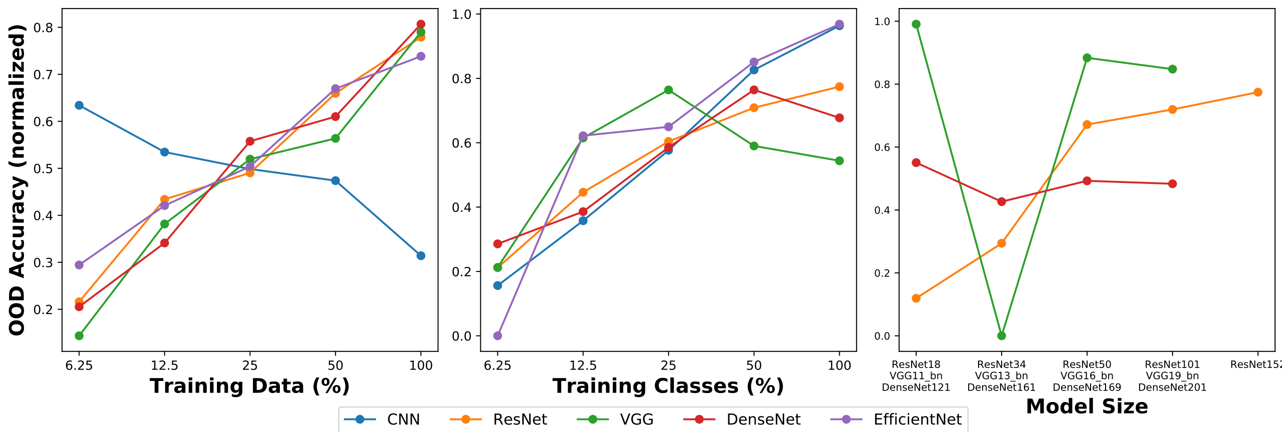

Figure 3. OOD accuracy averaged across all 70 train-target pairs in our study when scaling three parameters: amount of training data (left), number of training classes (center) and model size (right). The first two x-axis are a percentage of the maximum value (per dataset) and the third lists the architecture variants. Y-axis is the 5-way 1-shot normalized accuracy. Performance is normalized to show trend. For unnormalized performance, including power laws, see Figure 4.

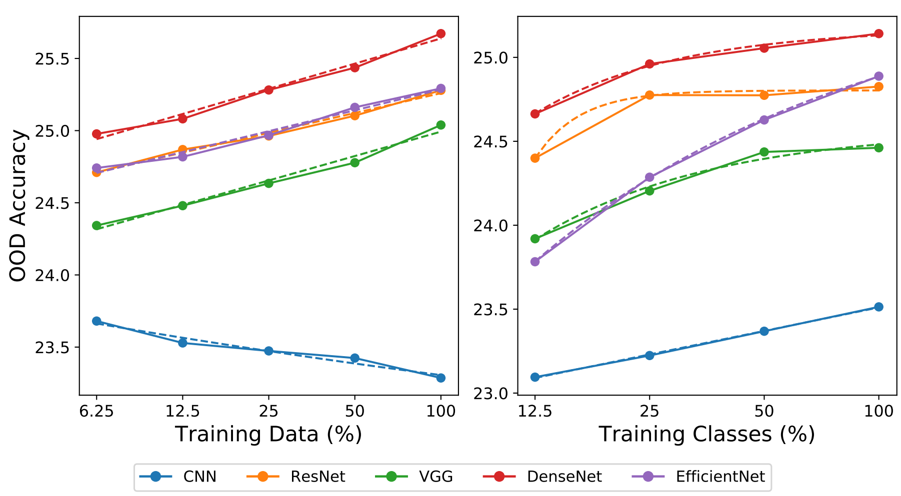

Figure 4. OOD accuracy averaged across all 70 train-target pairs, with power laws. (Left) scaling the amount of training data and (right) scaling the number of classes. X-axis are a percentage of the maximum value, per dataset. Y-axis is the 5-way 1-shot accuracy. Power laws are the dashed lines. See Table 1 for their constants.

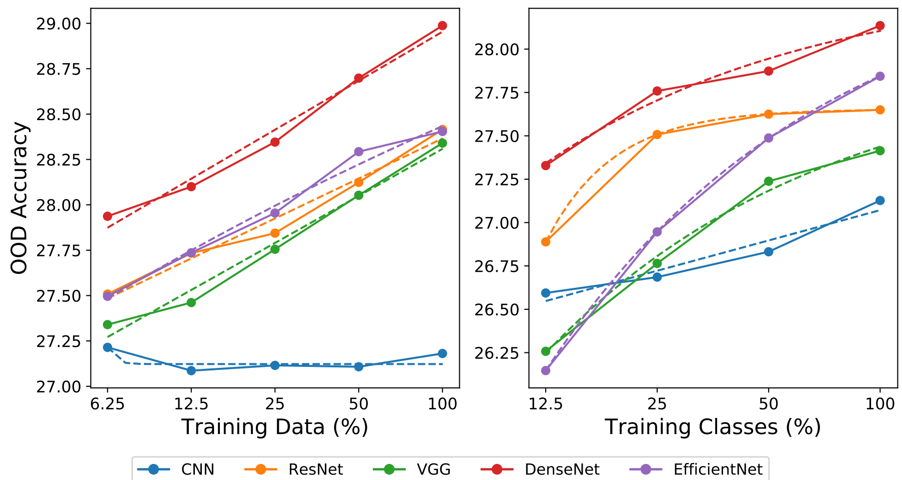

Figure 5. OOD accuracy averaged across all 70 train-target pairs, with power laws. (Left) scaling the amount of training data and (right) scaling the number of classes. X-axis are a percentage of the maximum value, per dataset. Y-axis is the 5-way 5-shot accuracy. For 5-way 1-shot accuracy, see section B. Power laws are the dashed lines. See Table 2 for their constants.

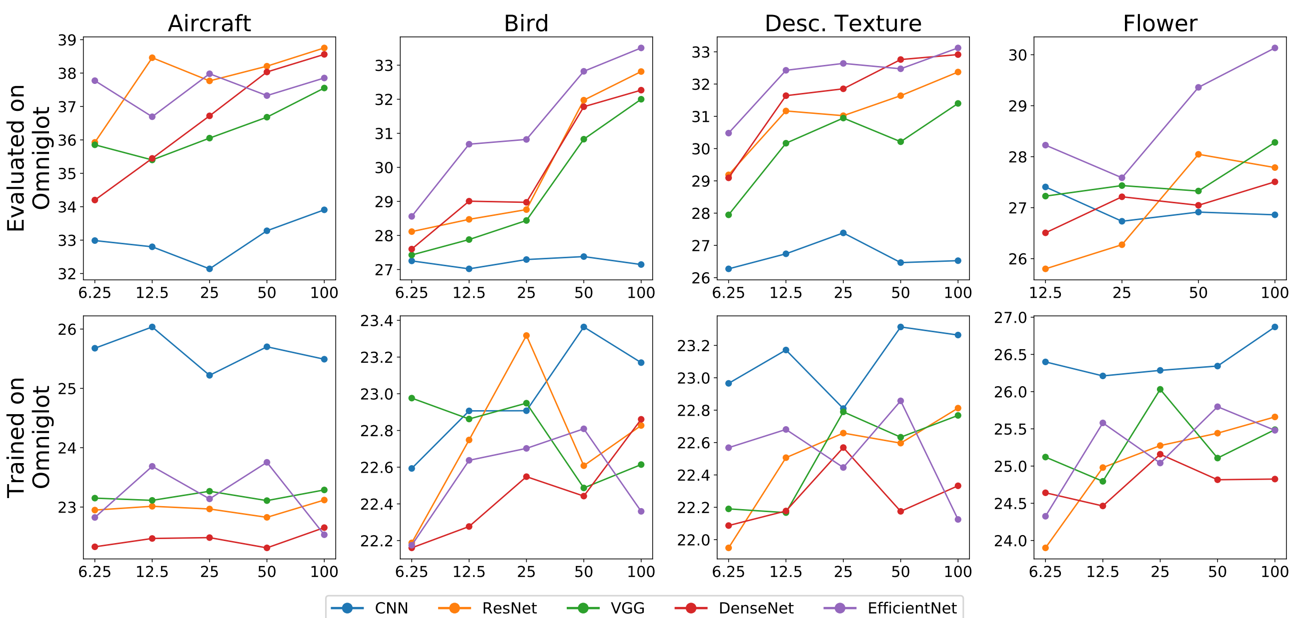

Figure 6. Omniglot results. (Top row) models trained on the dataset marked on top of each plot and evaluated on Omniglot. (Bottom row) models trained on Omniglot and evaluated on the dataset marked on top of each respective column. X-axis is the percentage of the total training data and y-axis is the 5-way 5-shot accuracy. For 5-way 1-shot accuracy, see section B.

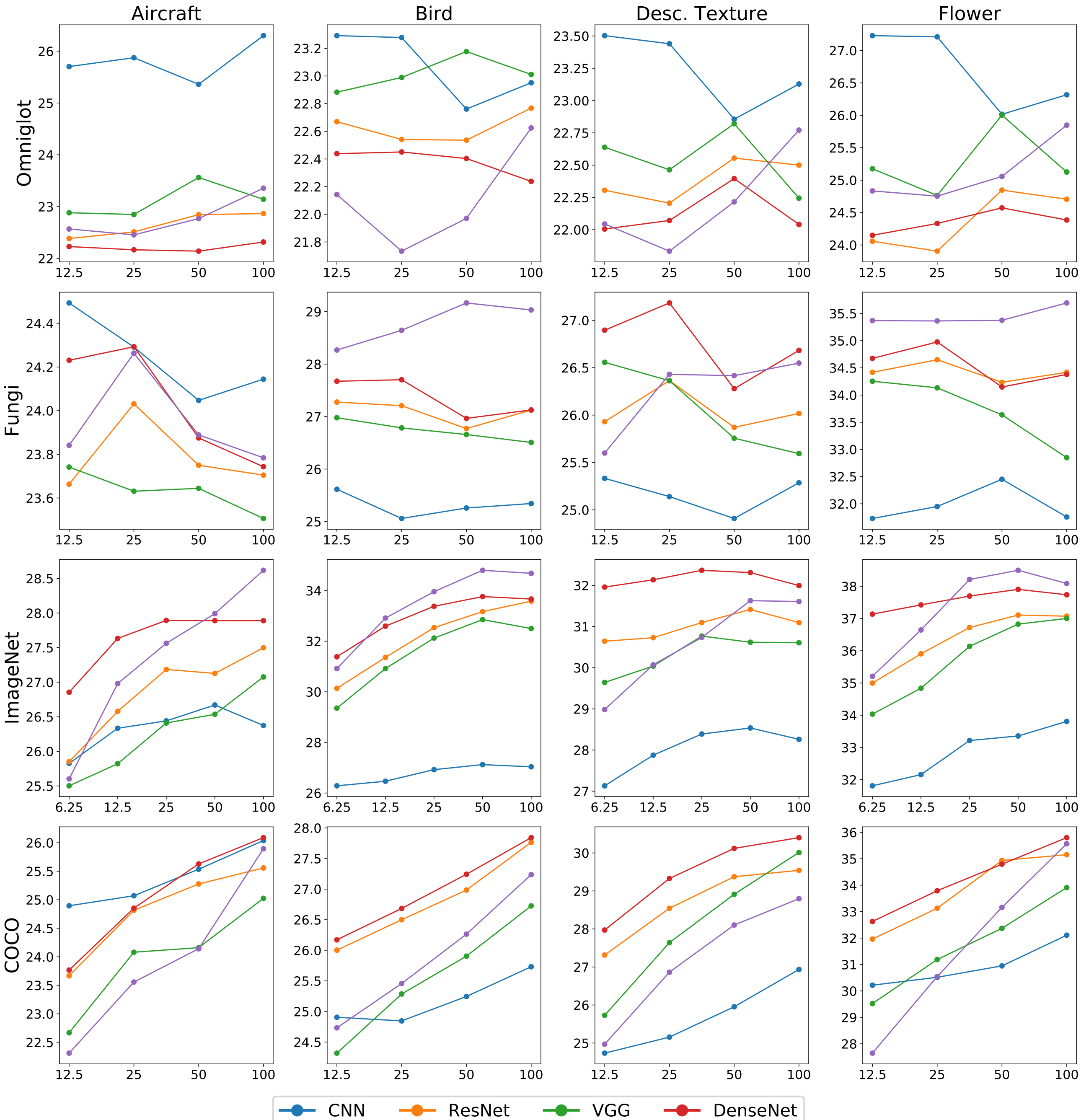

Figure 7. Scaling number of training classes results for various train-target pairs. For each plot, models are trained on the dataset marked on the left of each respective row and evaluated on the dataset marked top of each respective column. X-axis is the percentage of the total number of training classes of each dataset and y-axis is the 5-way 5-shot accuracy. For 5-way 1-shot results, see section B.

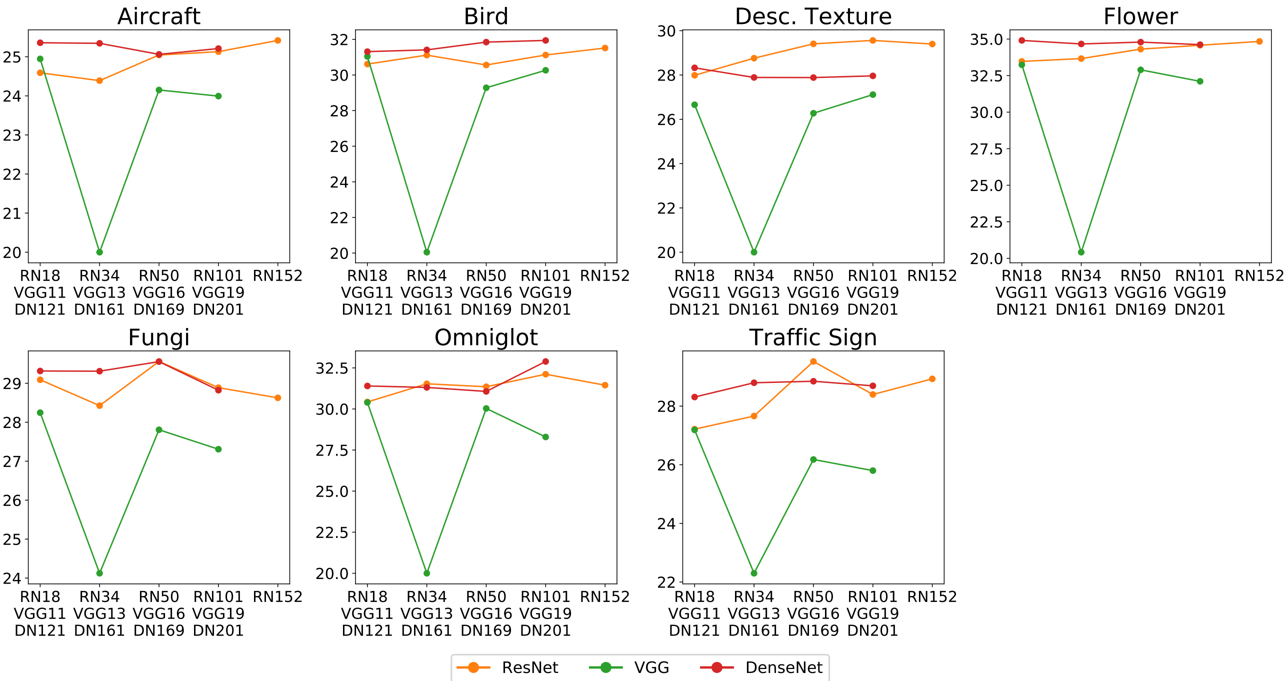

Figure 8. Scaling model size results. All models are trained on ImageNet and evaluated on the dataset marked top of each plot. X-axis lists each architecture variant with RN meaning ResNet and DN meaning DenseNet. Y-axis is the 5-way 1-shot accuracy. For 5-way 5-shot results, see Figure 9.

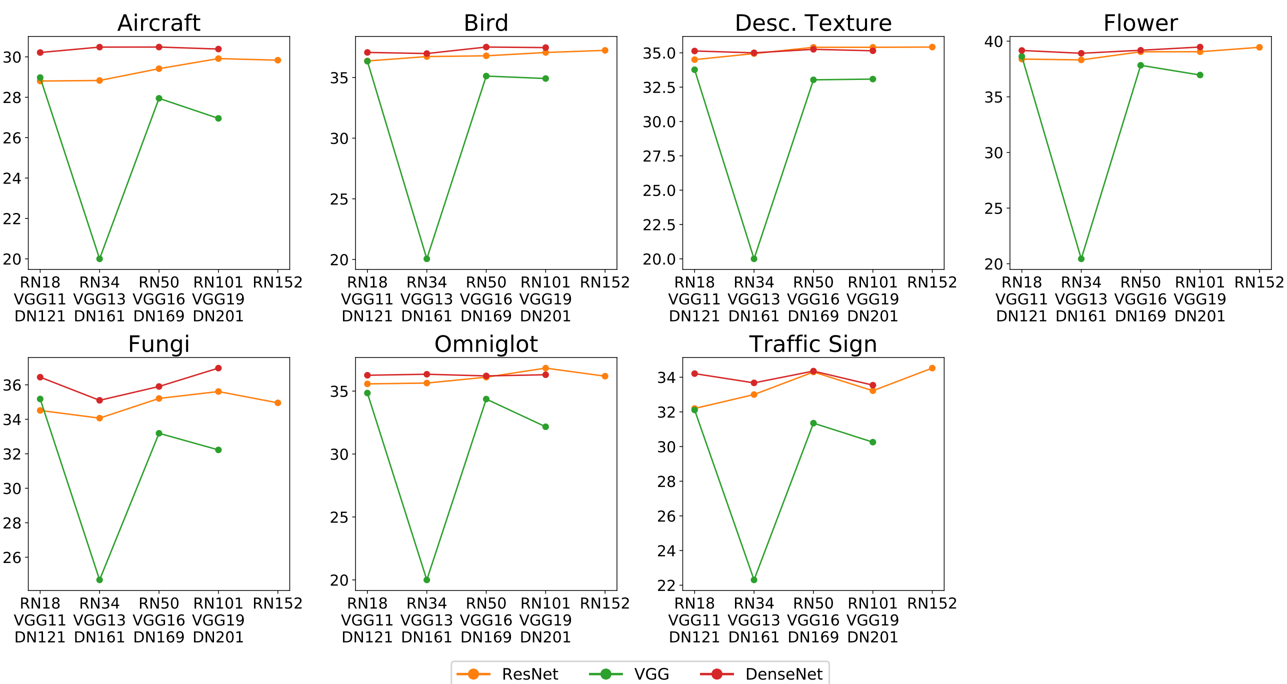

Figure 9. Scaling model size results. All models are trained on ImageNet and evaluated on the dataset marked top of each plot. X-axis lists each architecture variant with RN meaning ResNet and DN meaning DenseNet. Y-axis is the 5-way 5-shot accuracy. For 5-way 1-shot results, see Figure 8.

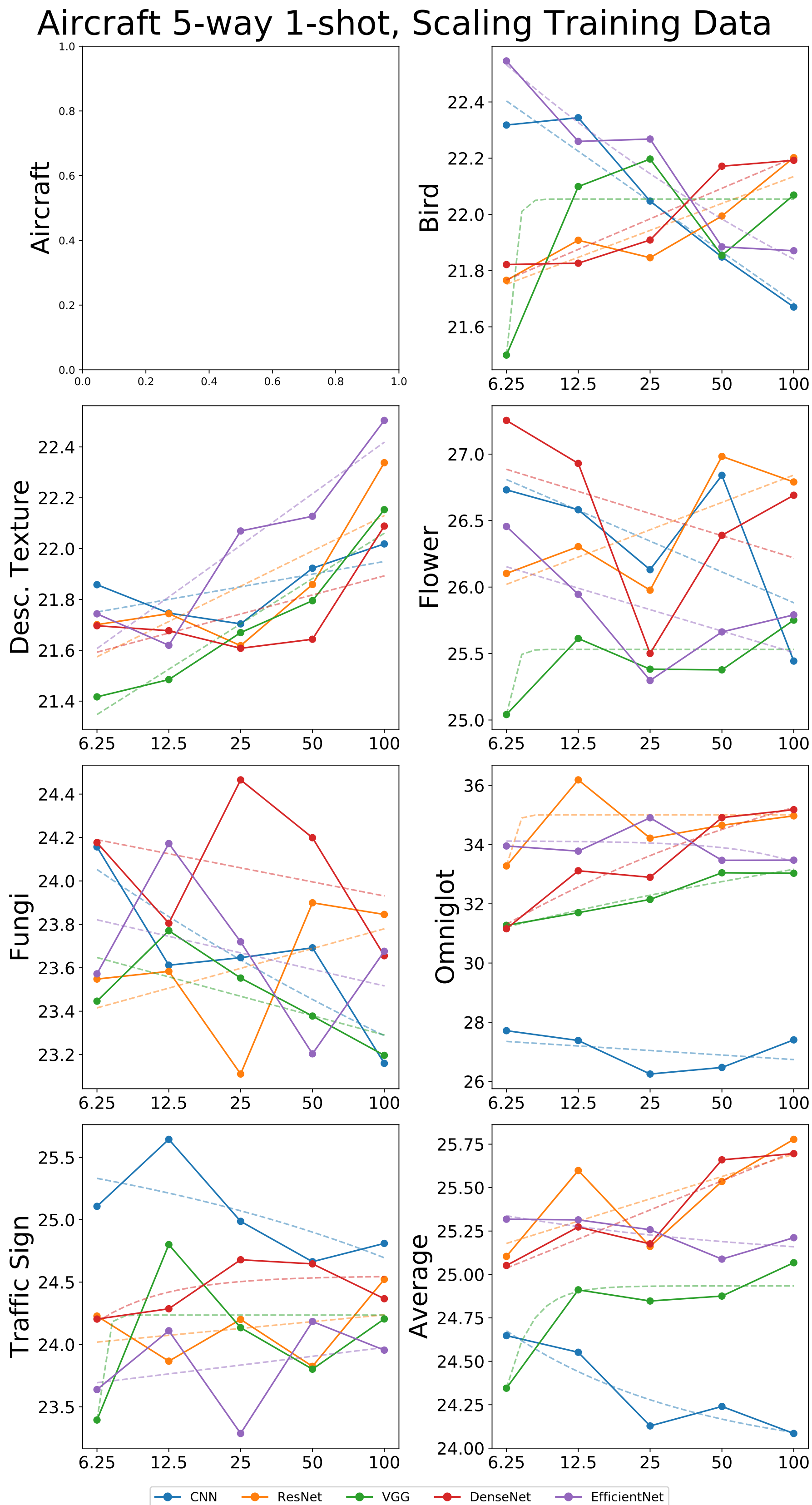

Figure 10. Scaling training data 5-way 1-shot results for models trained on Aircraft. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 1-shot accuracy.

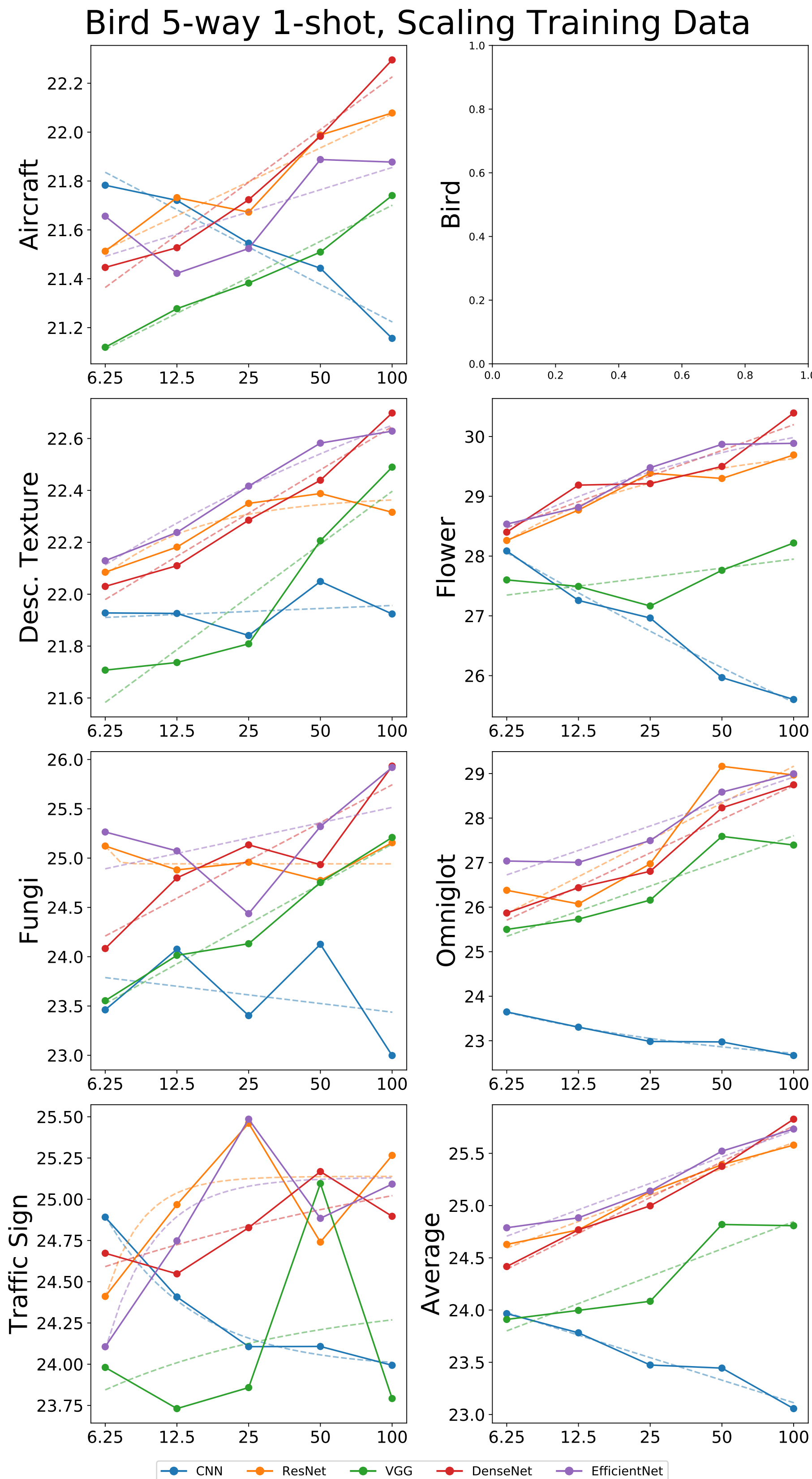

Figure 11. Scaling training data 5-way 1-shot results for models trained on Bird. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 1-shot accuracy.

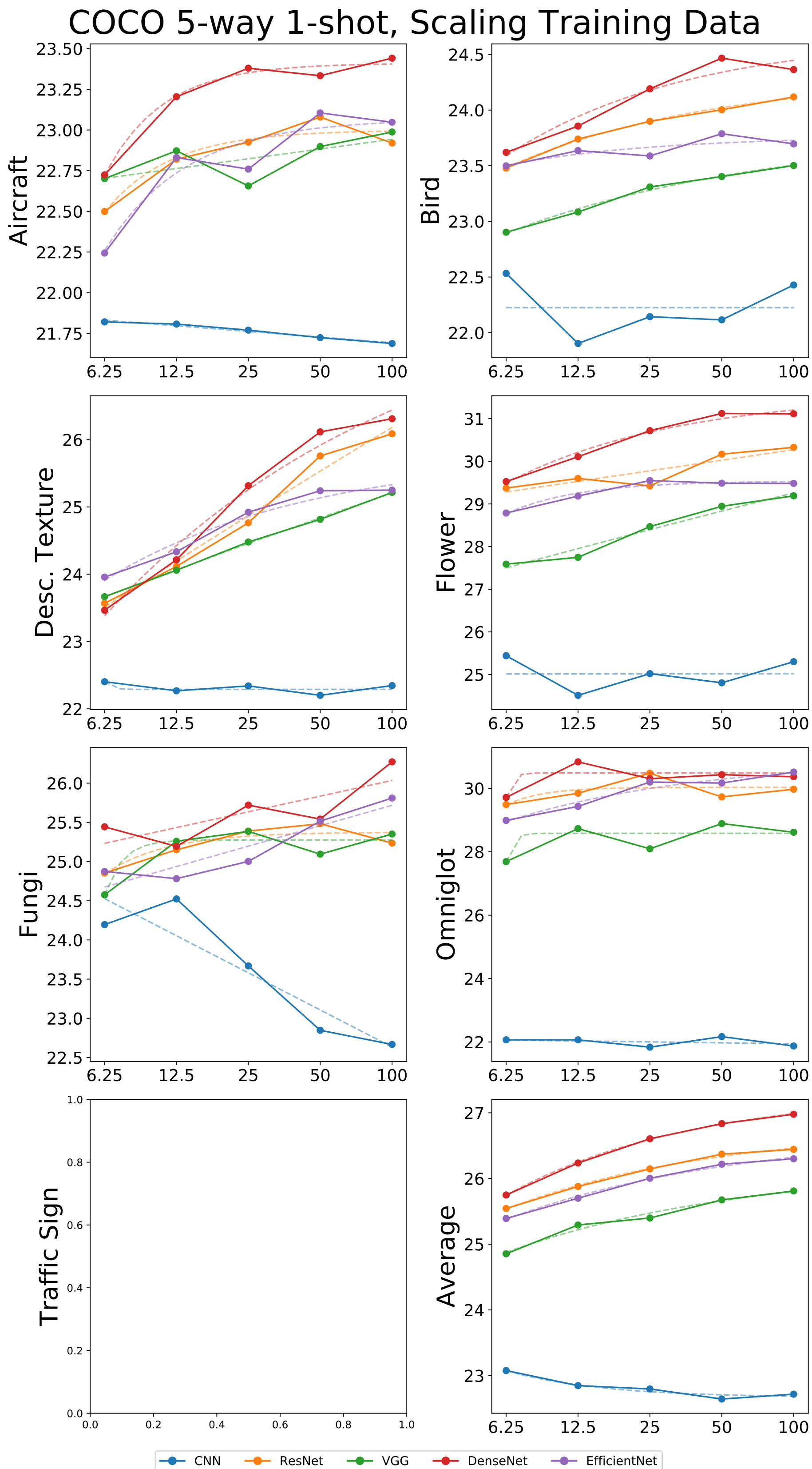

Figure 12. Scaling training data 5-way 1-shot results for models trained on COCO. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 1-shot accuracy.

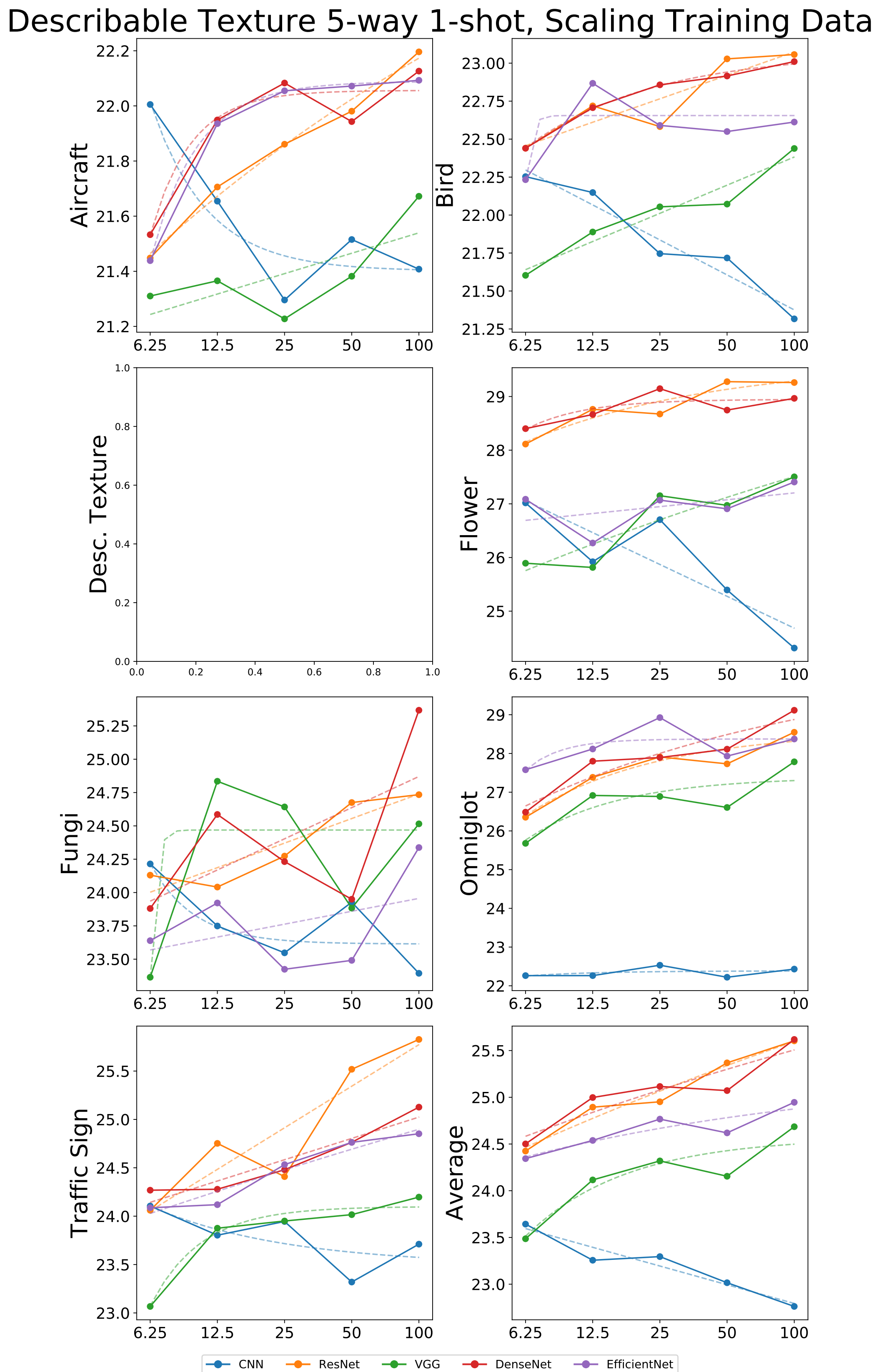

Figure 13. Scaling training data 5-way 1-shot results for models trained on Describable Texture. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 1-shot accuracy.

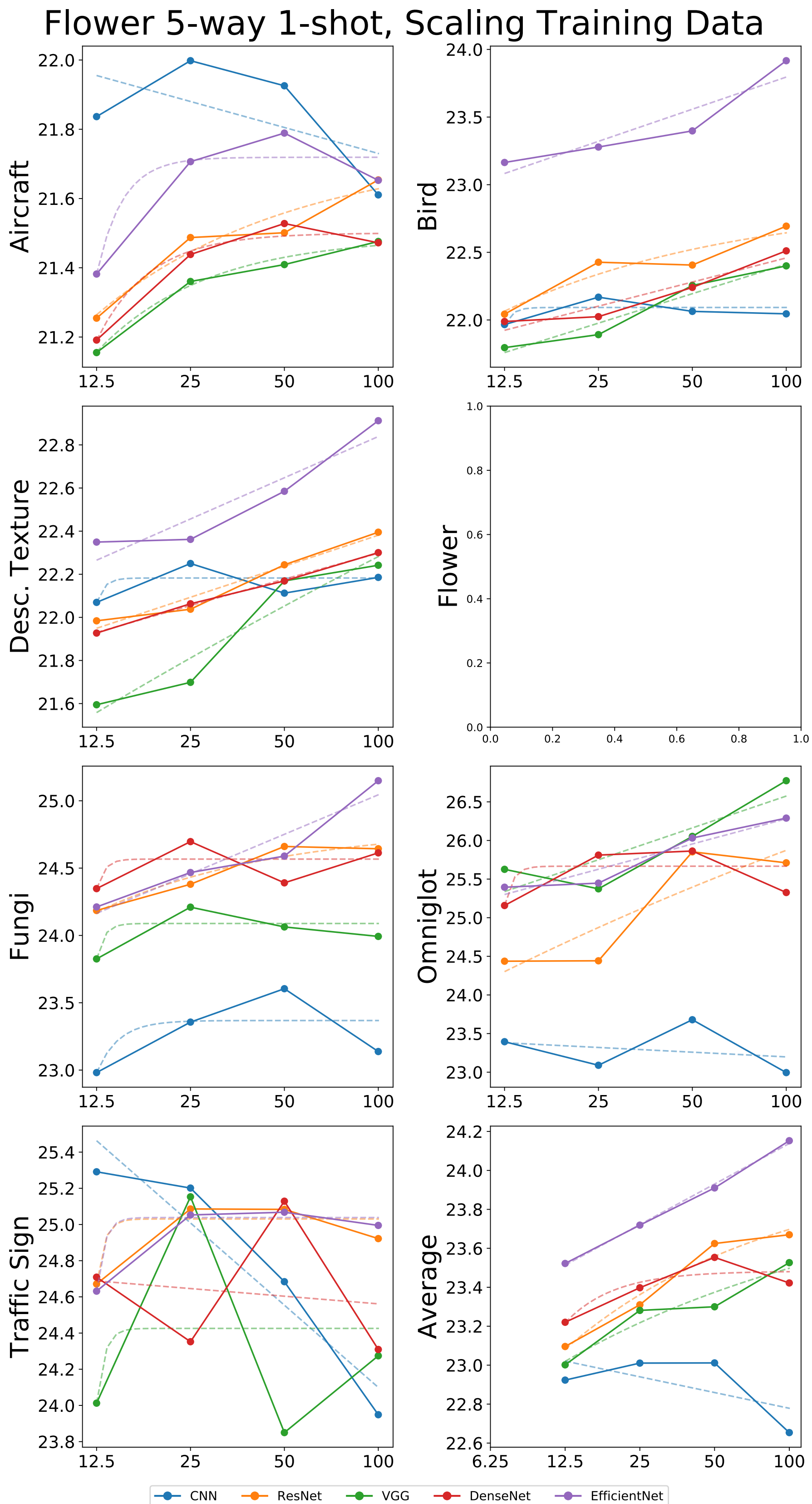

Figure 14. Scaling training data 5-way 1-shot results for models trained on Flower. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 1-shot accuracy.

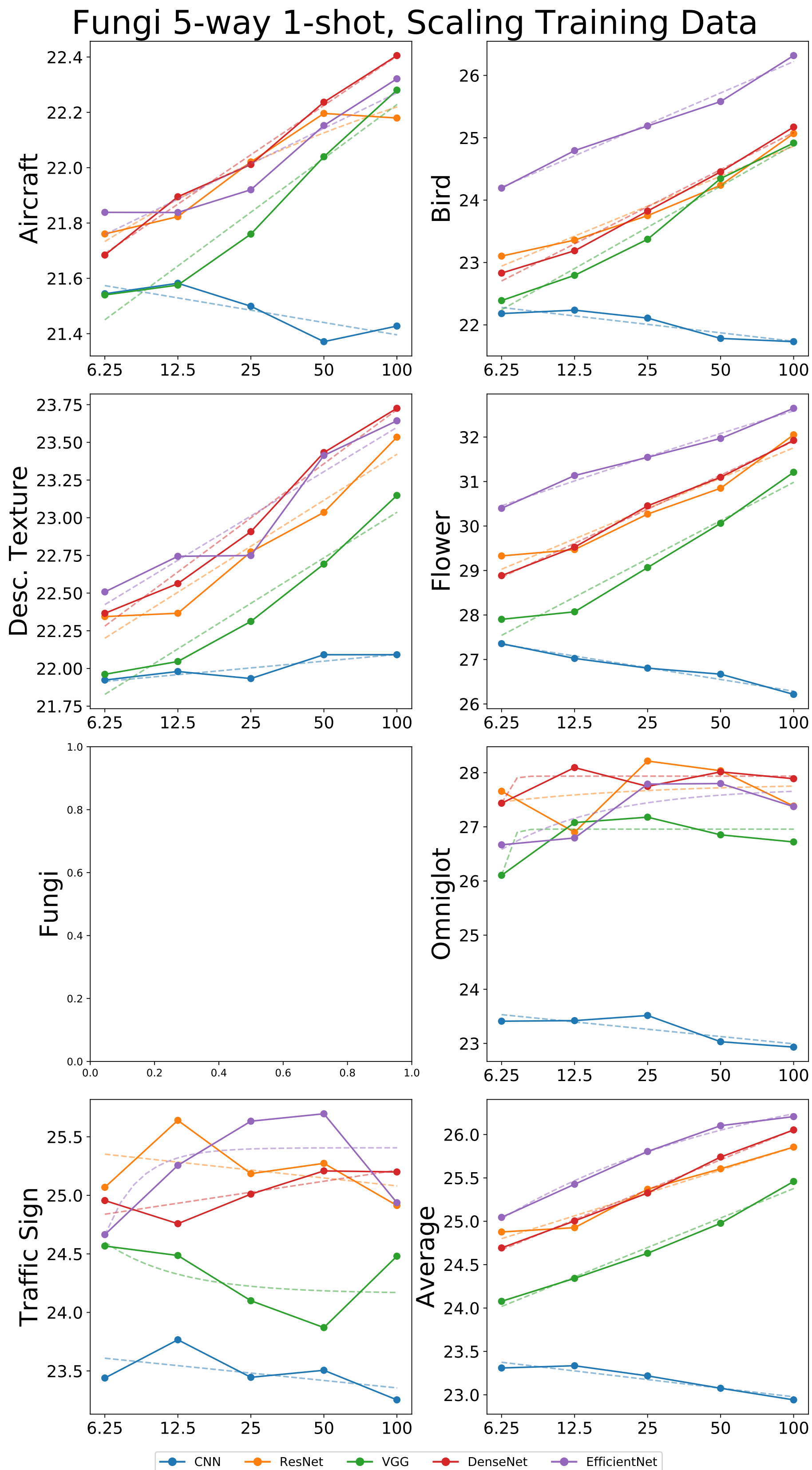

Figure 15. Scaling training data 5-way 1-shot results for models trained on Fungi. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 1-shot accuracy.

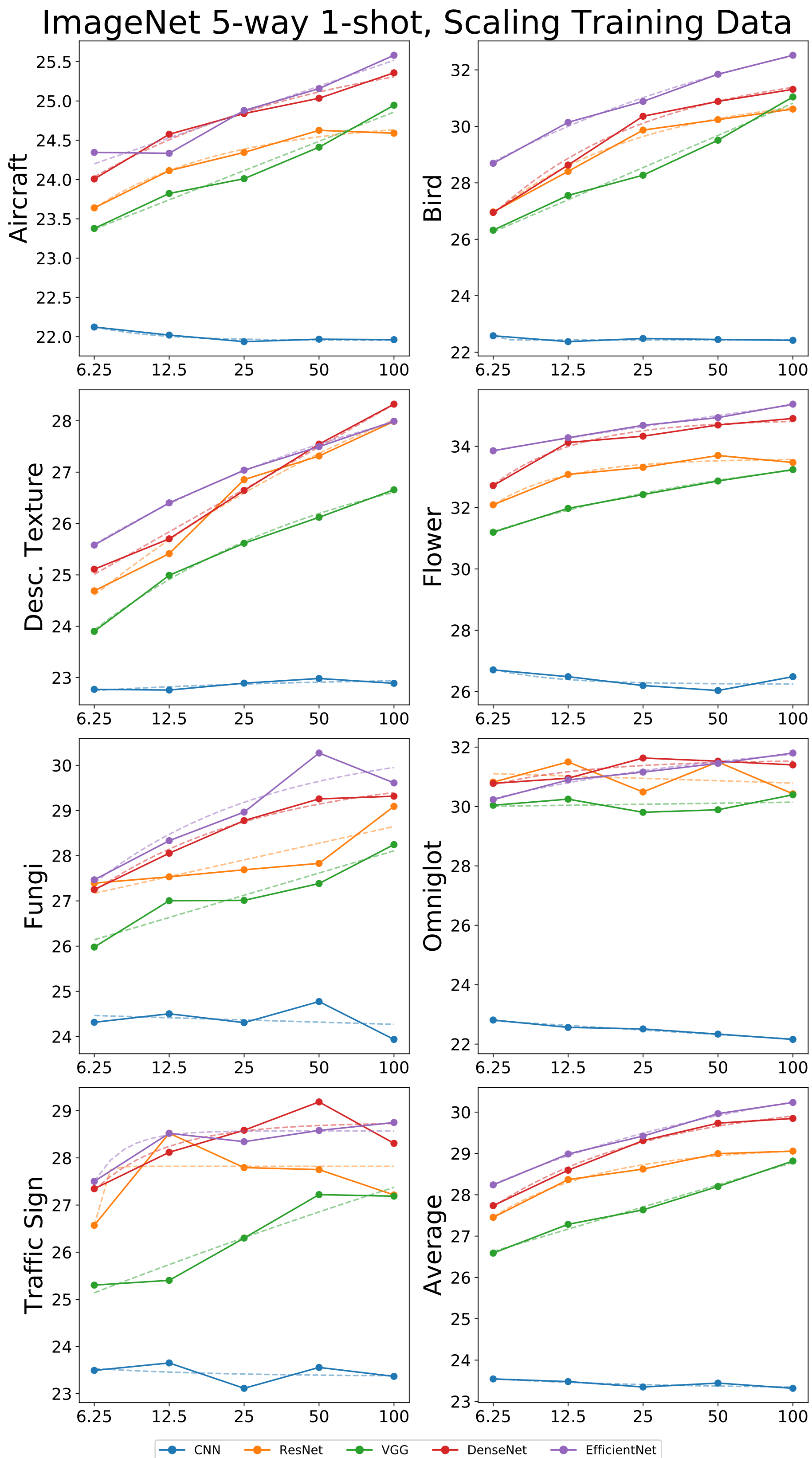

Figure 16. Scaling training data 5-way 1-shot results for models trained on ImageNet. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 1-shot accuracy.

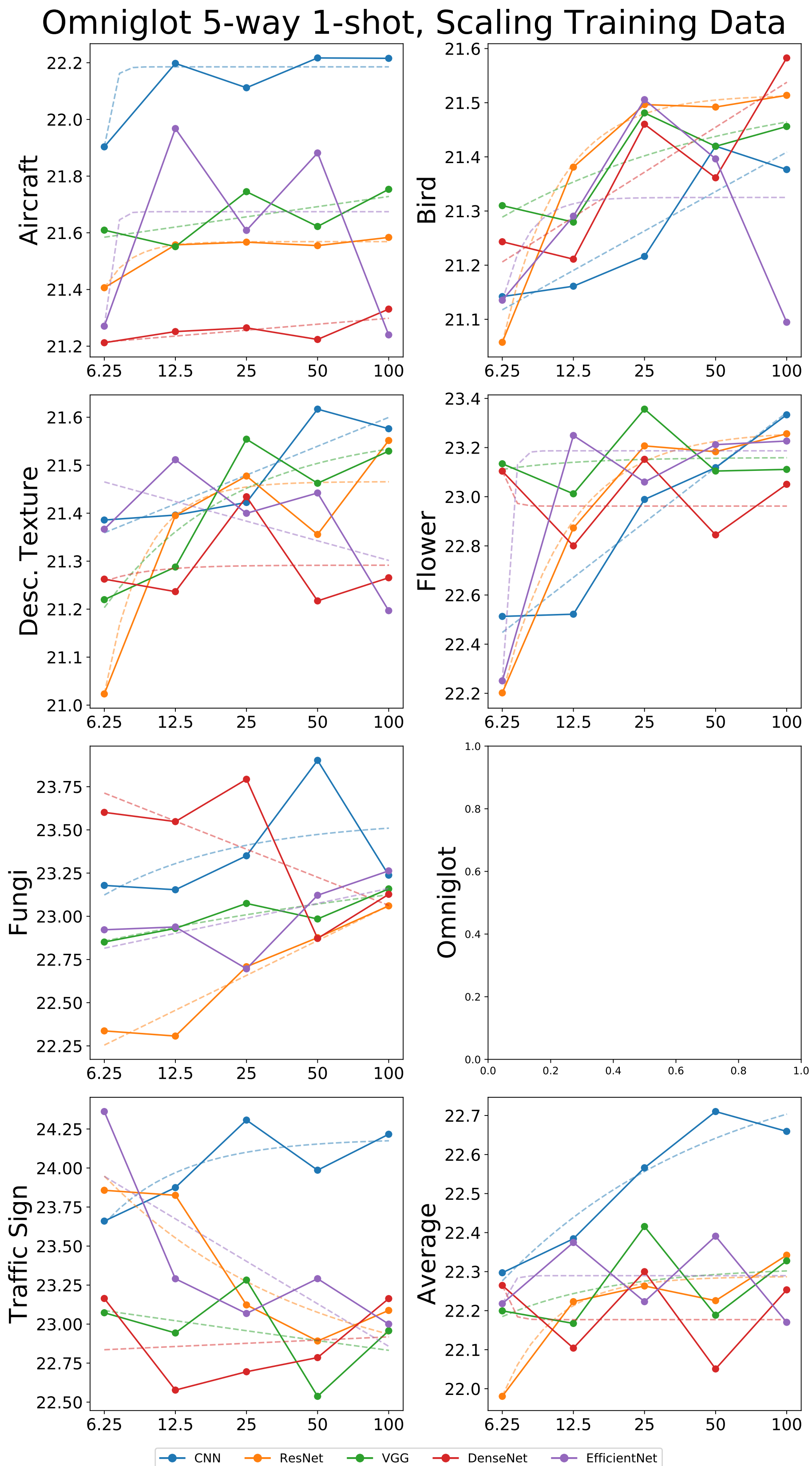

Figure 17. Scaling training data 5-way 1-shot results for models trained on Omniglot. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 1-shot accuracy.

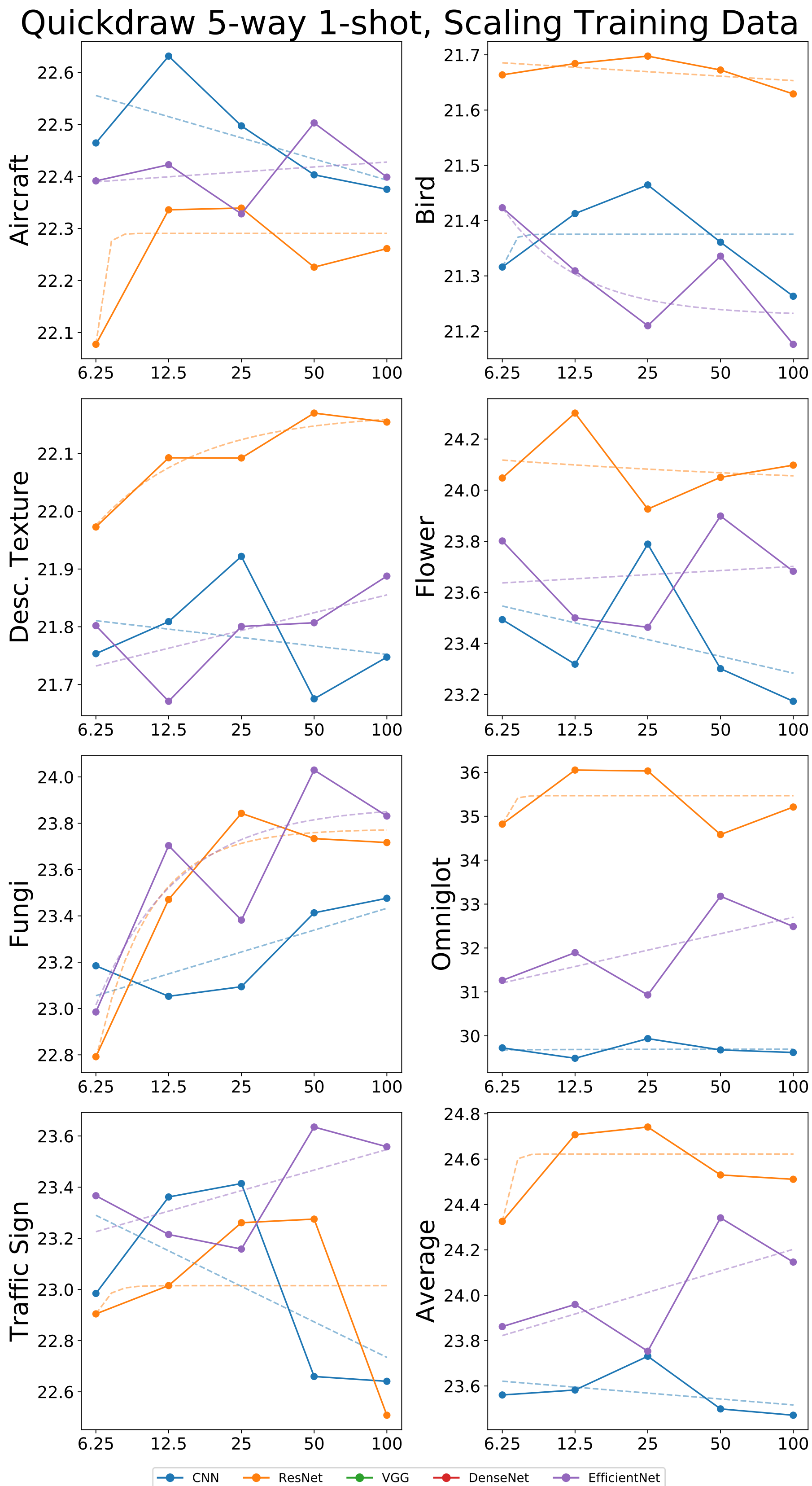

Figure 18. Scaling training data 5-way 1-shot results for models trained on Quickdraw. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 1-shot accuracy.

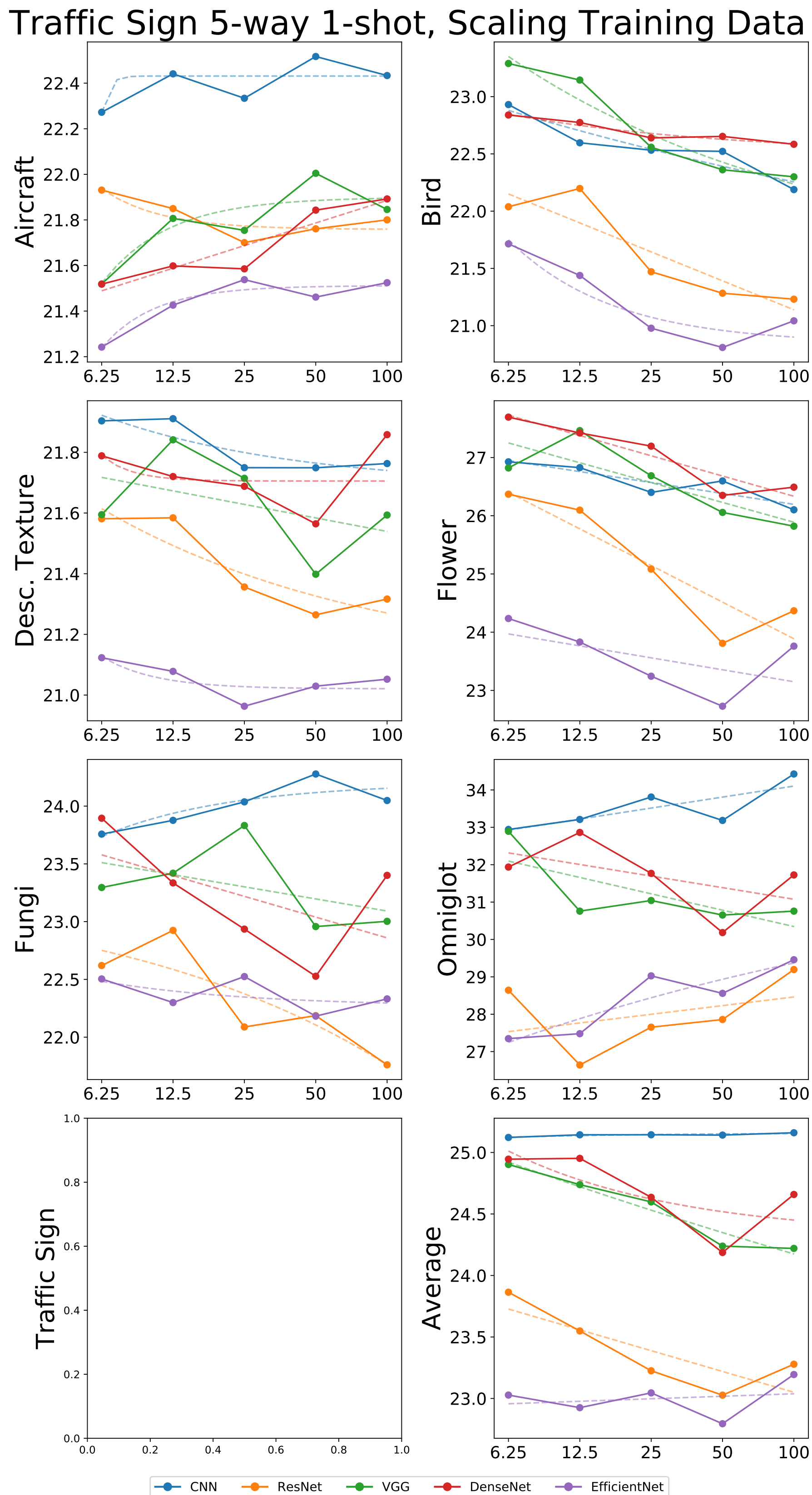

Figure 19. Scaling training data 5-way 1-shot results for models trained on Traffic Sign. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 1-shot accuracy.

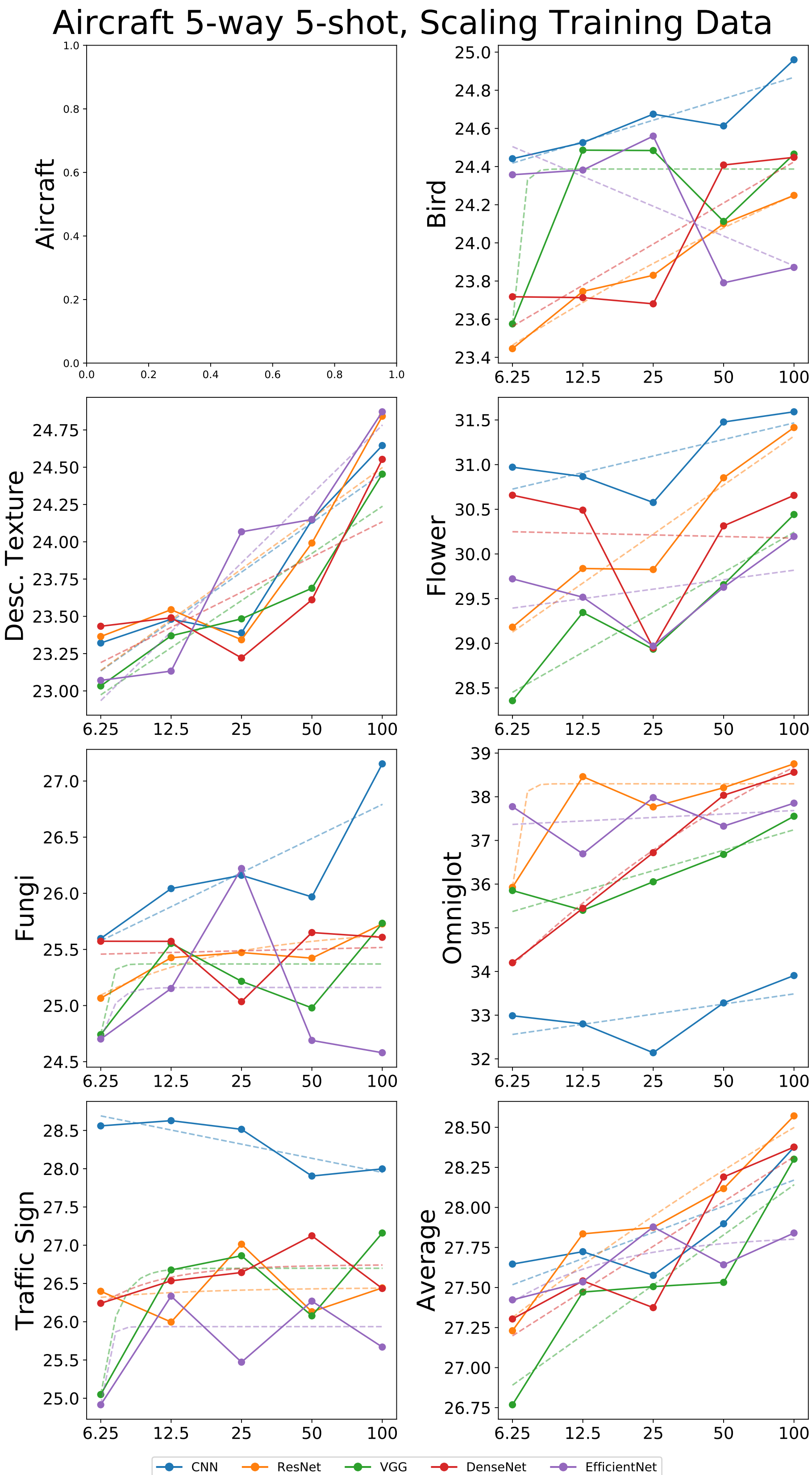

Figure 20. Scaling training data 5-way 5-shot results for models trained on Aircraft. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 5-shot accuracy.

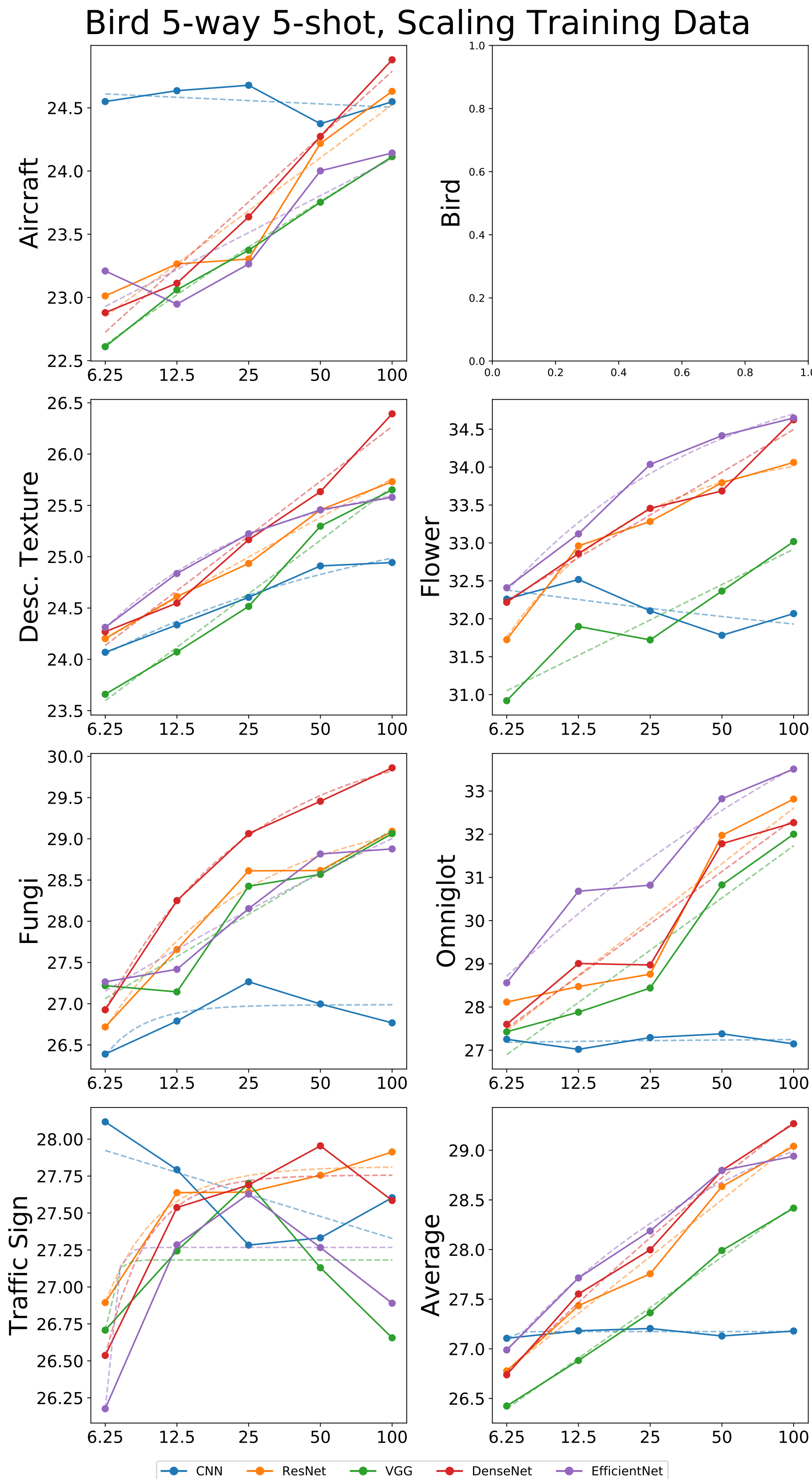

Figure 21. Scaling training data 5-way 5-shot results for models trained on Bird. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 5-shot accuracy.

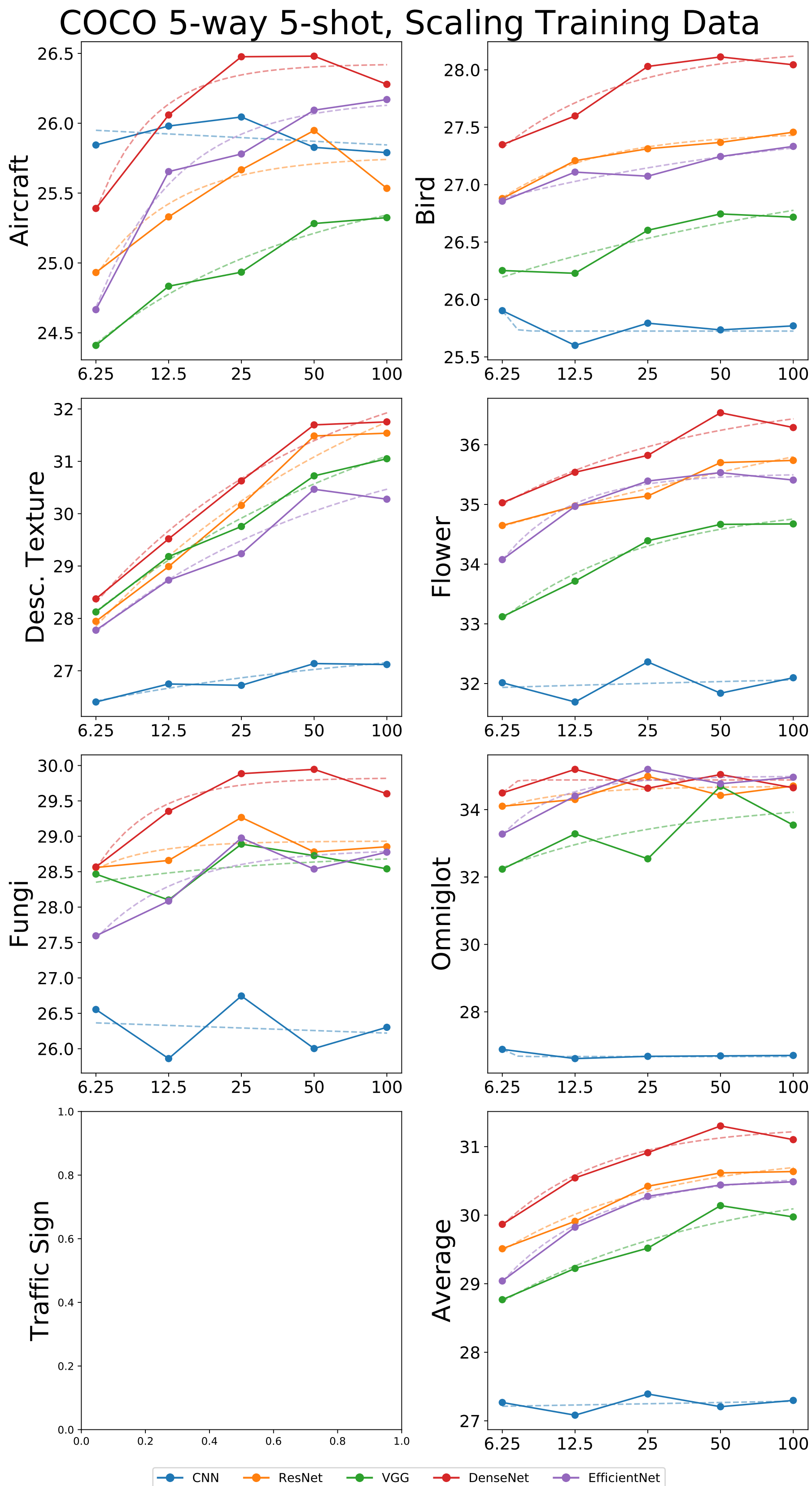

Figure 22. Scaling training data 5-way 5-shot results for models trained on COCO. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 5-shot accuracy.

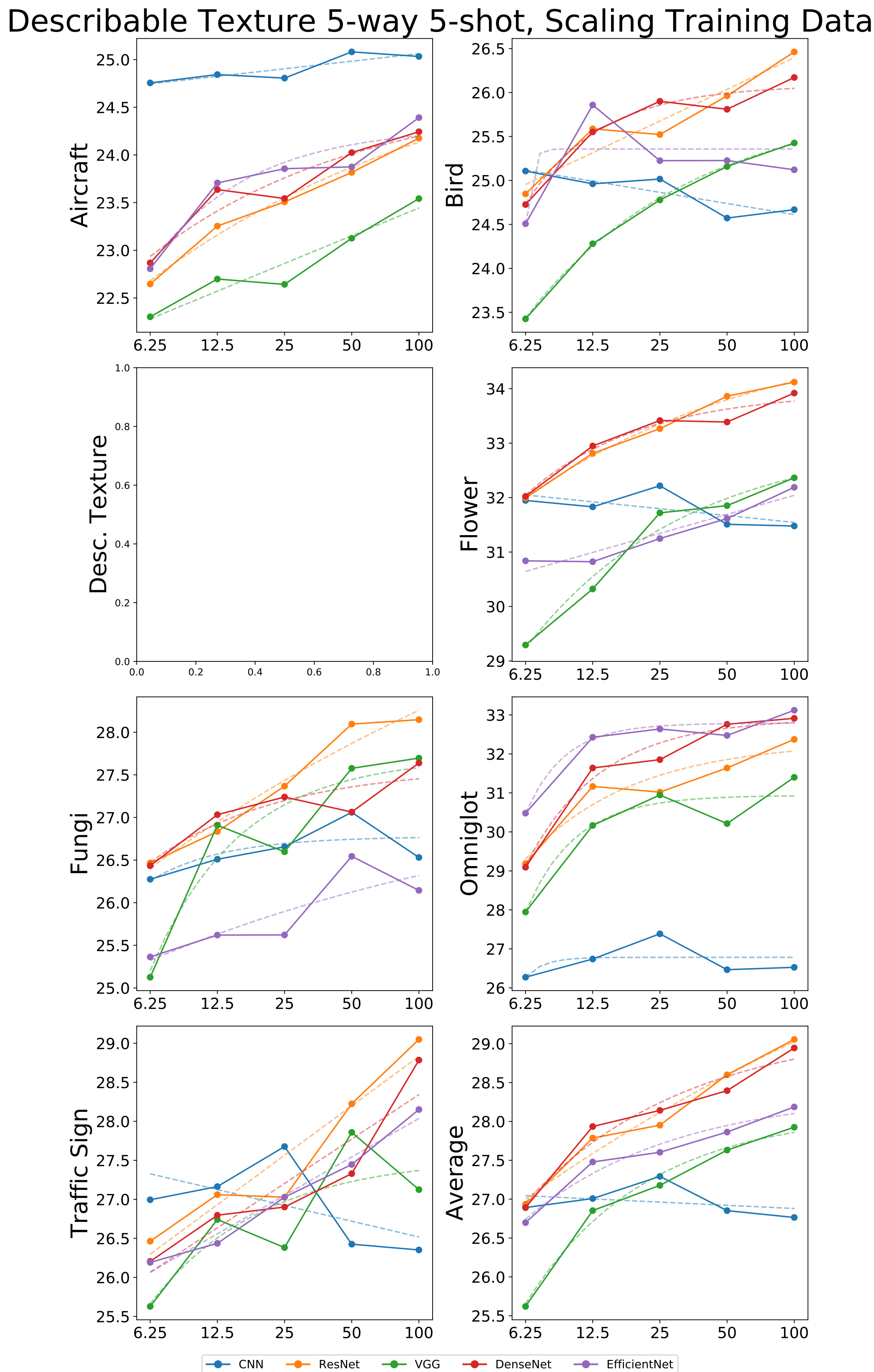

Figure 23. Scaling training data 5-way 5-shot results for models trained on Describable Texture. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 5-shot accuracy.

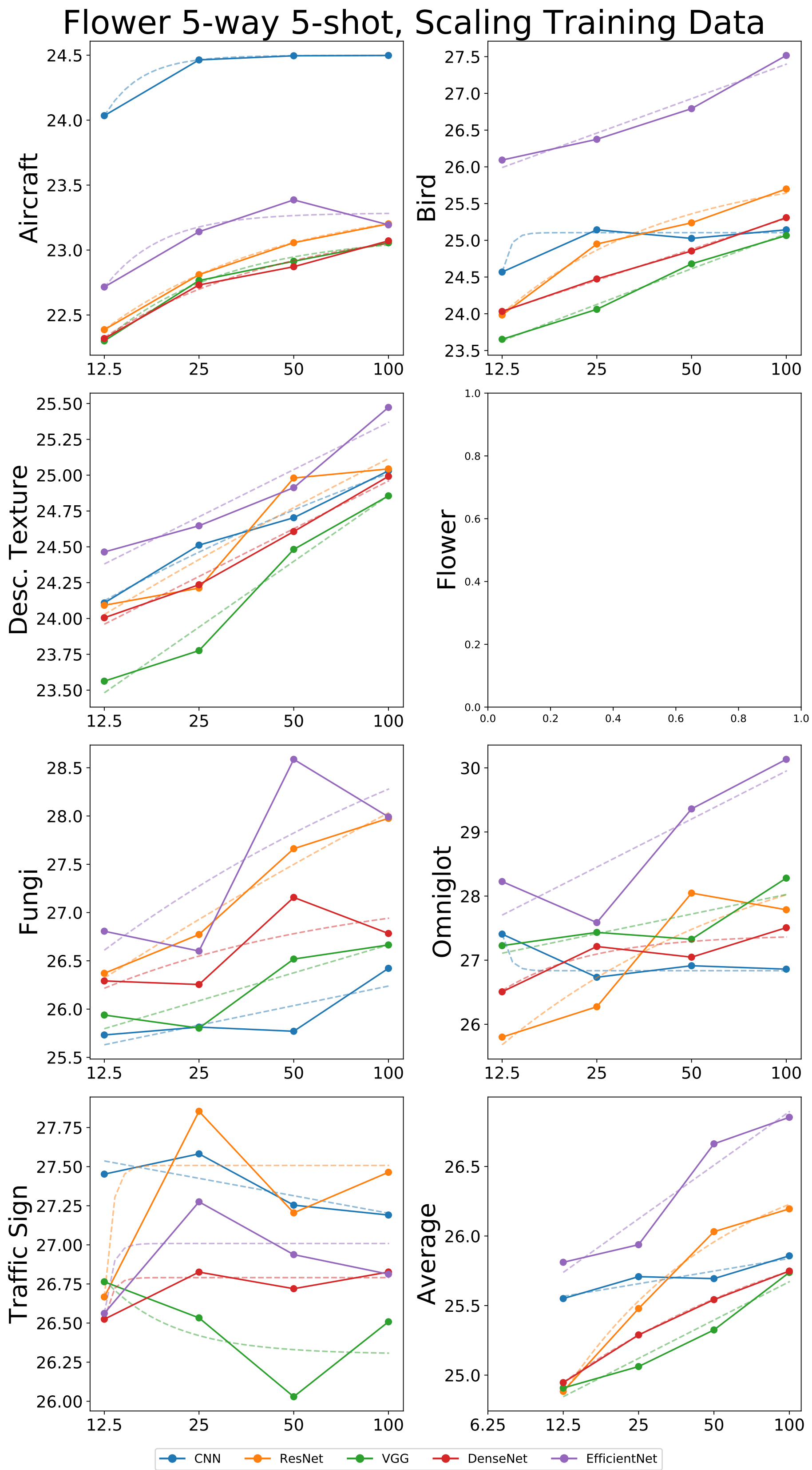

Figure 24. Scaling training data 5-way 5-shot results for models trained on Flower. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 5-shot accuracy.

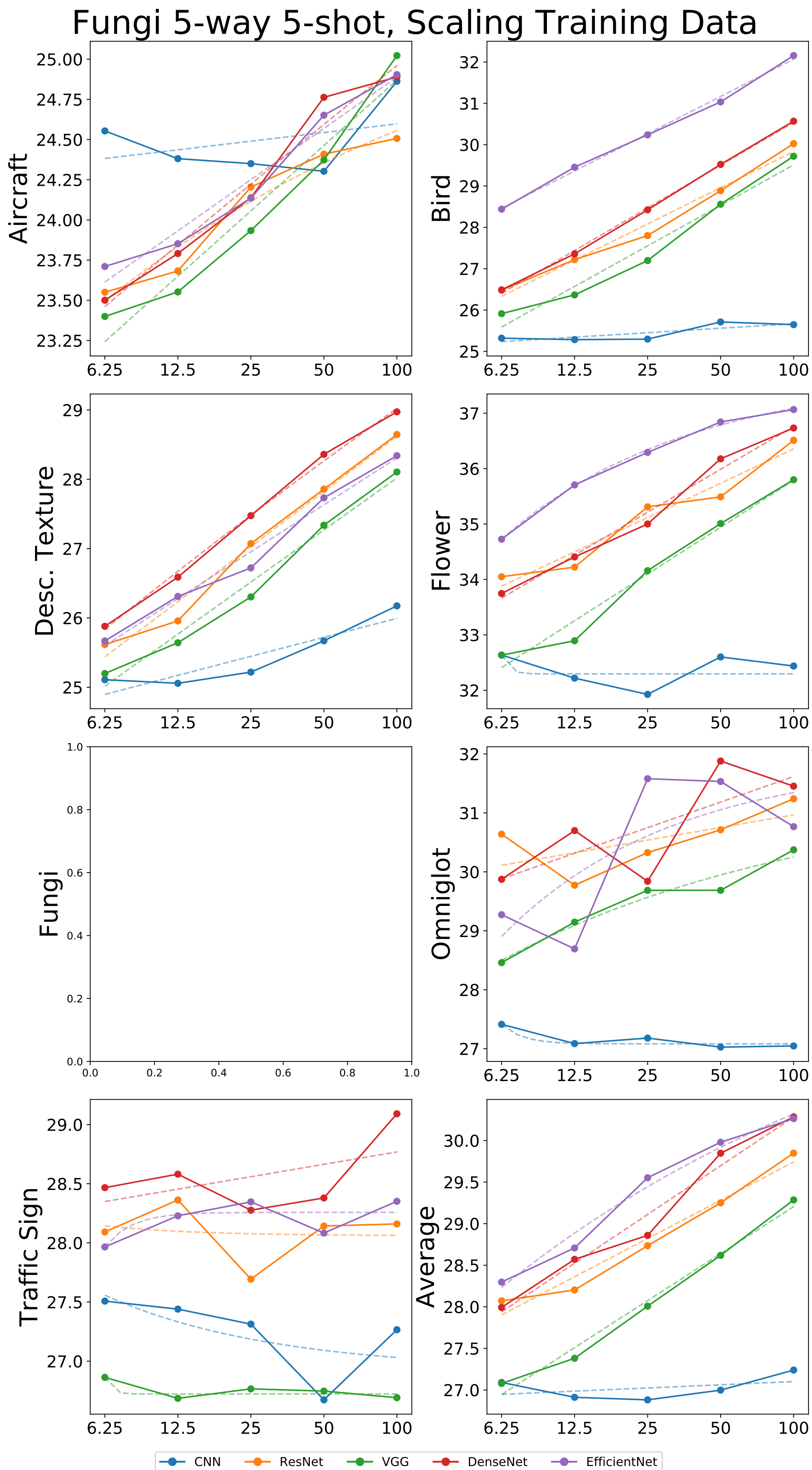

Figure 25. Scaling training data 5-way 5-shot results for models trained on Fungi. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 5-shot accuracy.

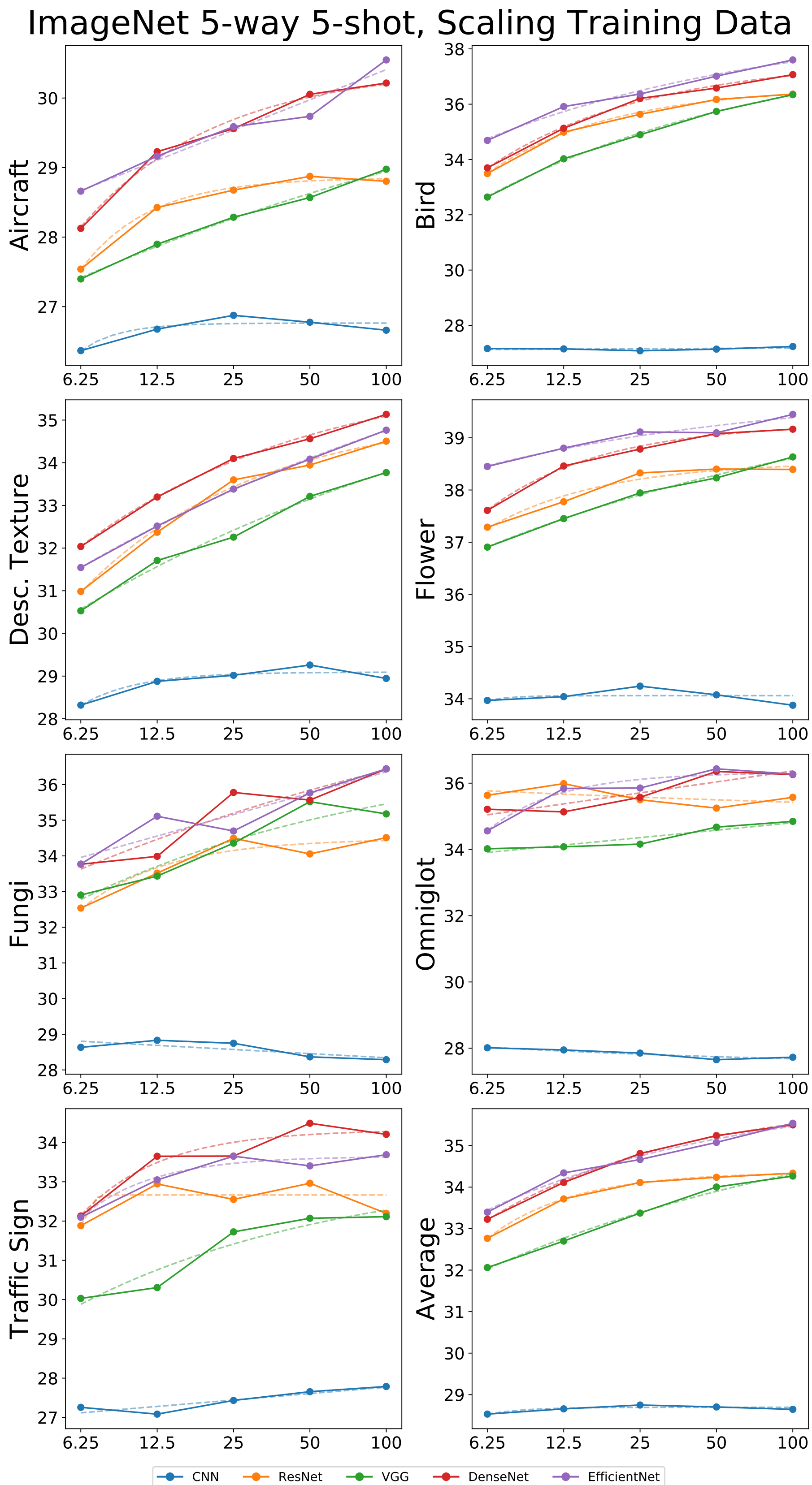

Figure 26. Scaling training data 5-way 5-shot results for models trained on ImageNet. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 5-shot accuracy.

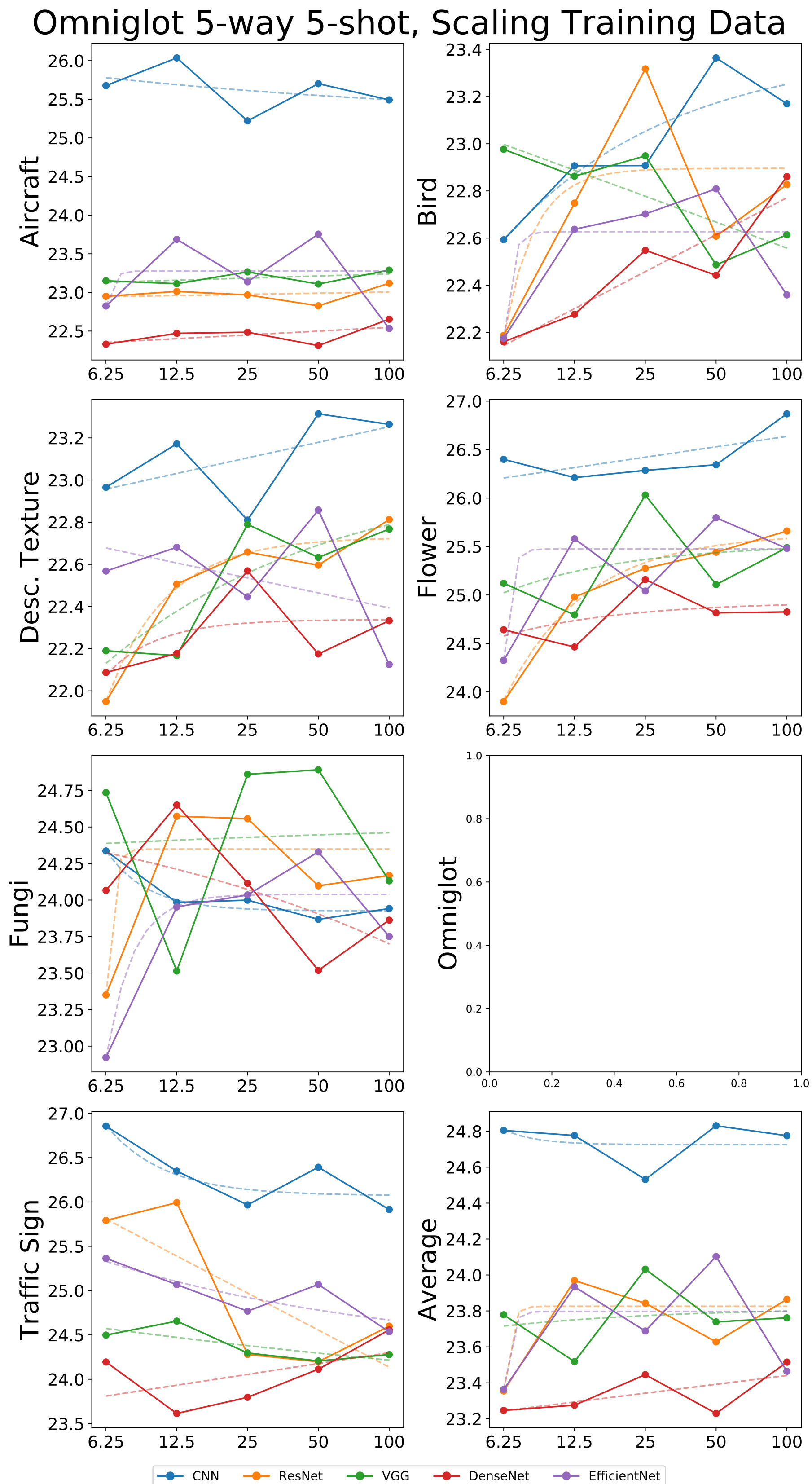

Figure 27. Scaling training data 5-way 5-shot results for models trained on Omniglot. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 5-shot accuracy.

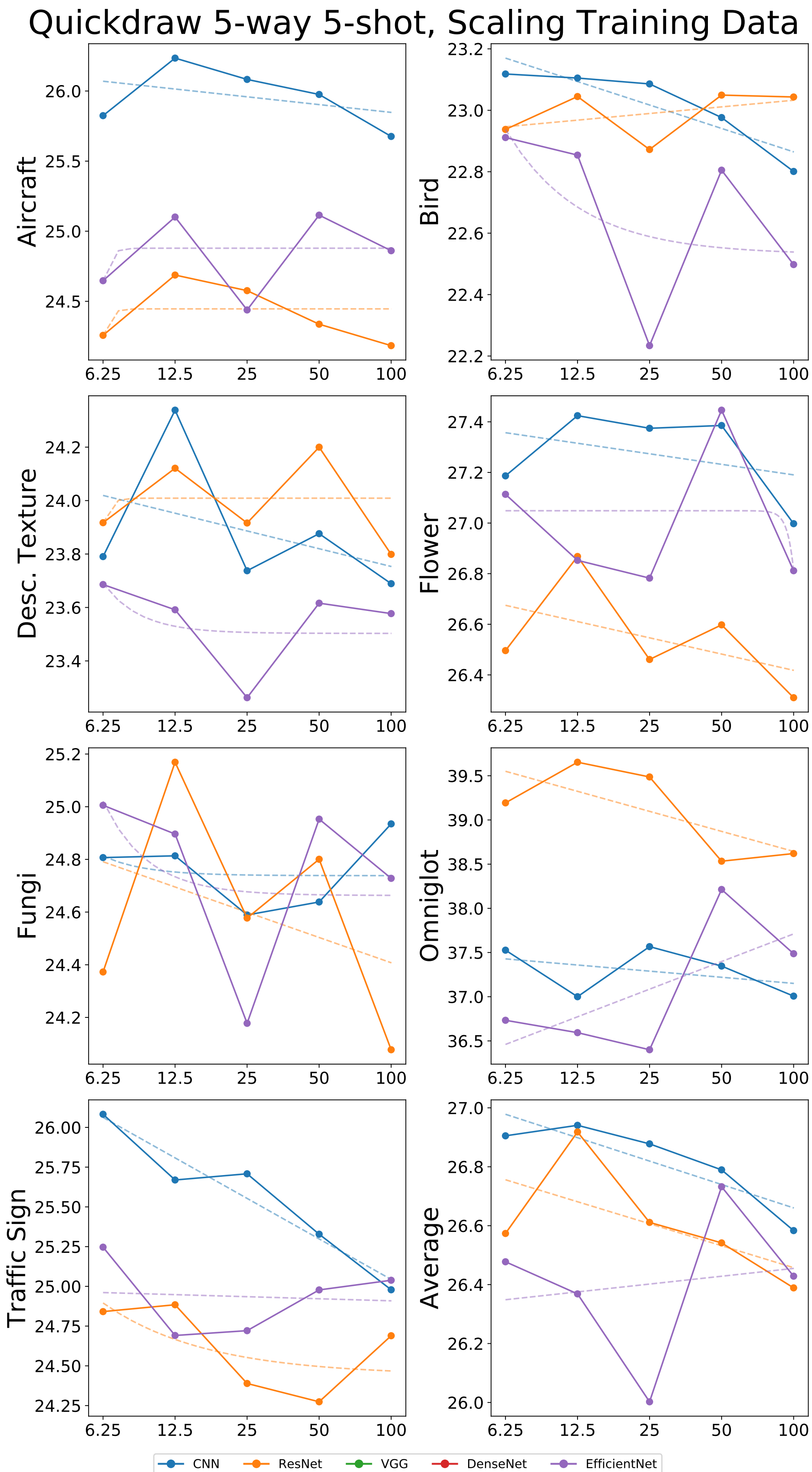

Figure 28. Scaling training data 5-way 5-shot results for models trained on Quickdraw. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 5-shot accuracy.

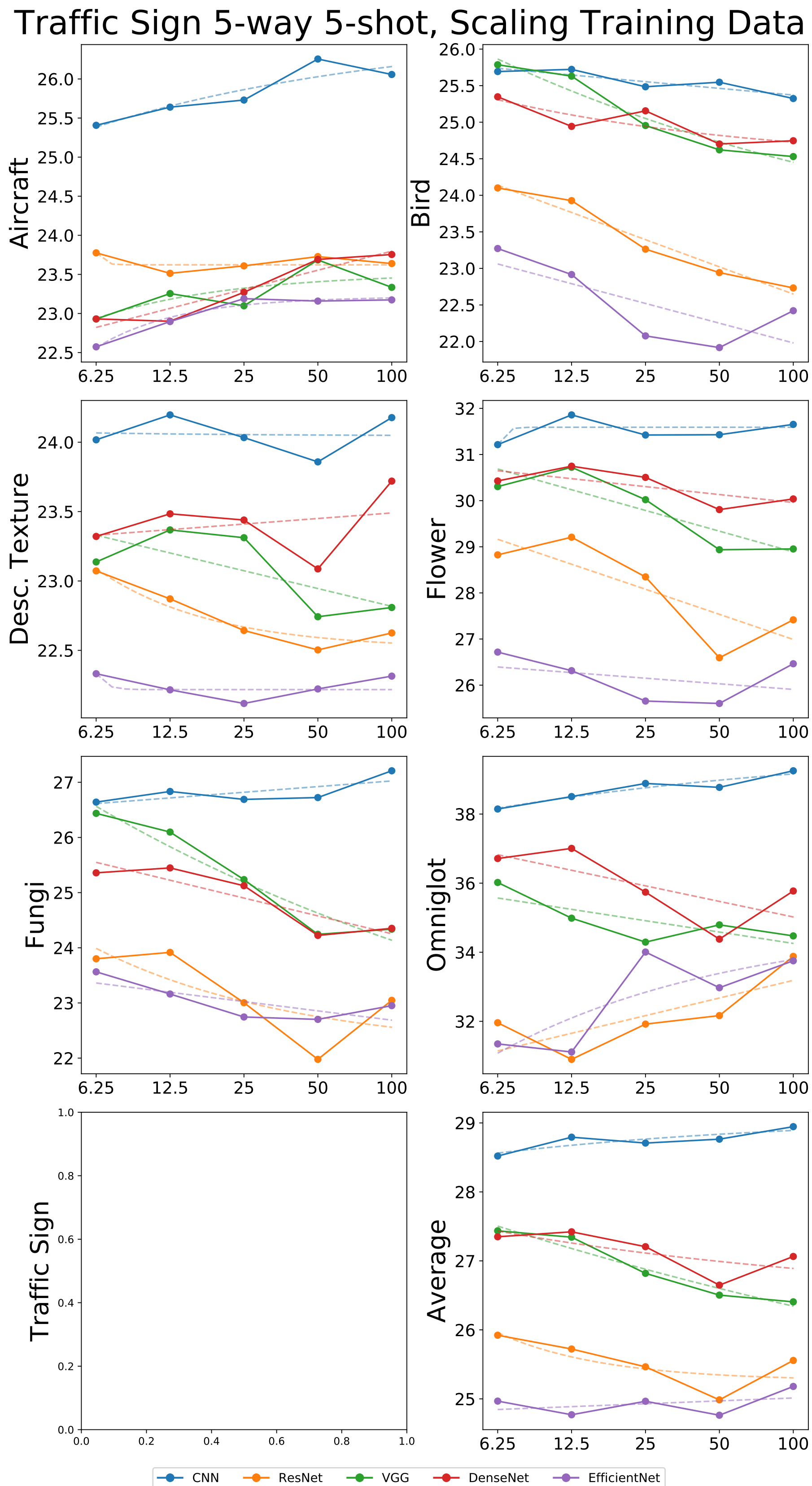

Figure 29. Scaling training data 5-way 5-shot results for models trained on Traffic Sign. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training data and y-axis is the 5-way 5-shot accuracy.

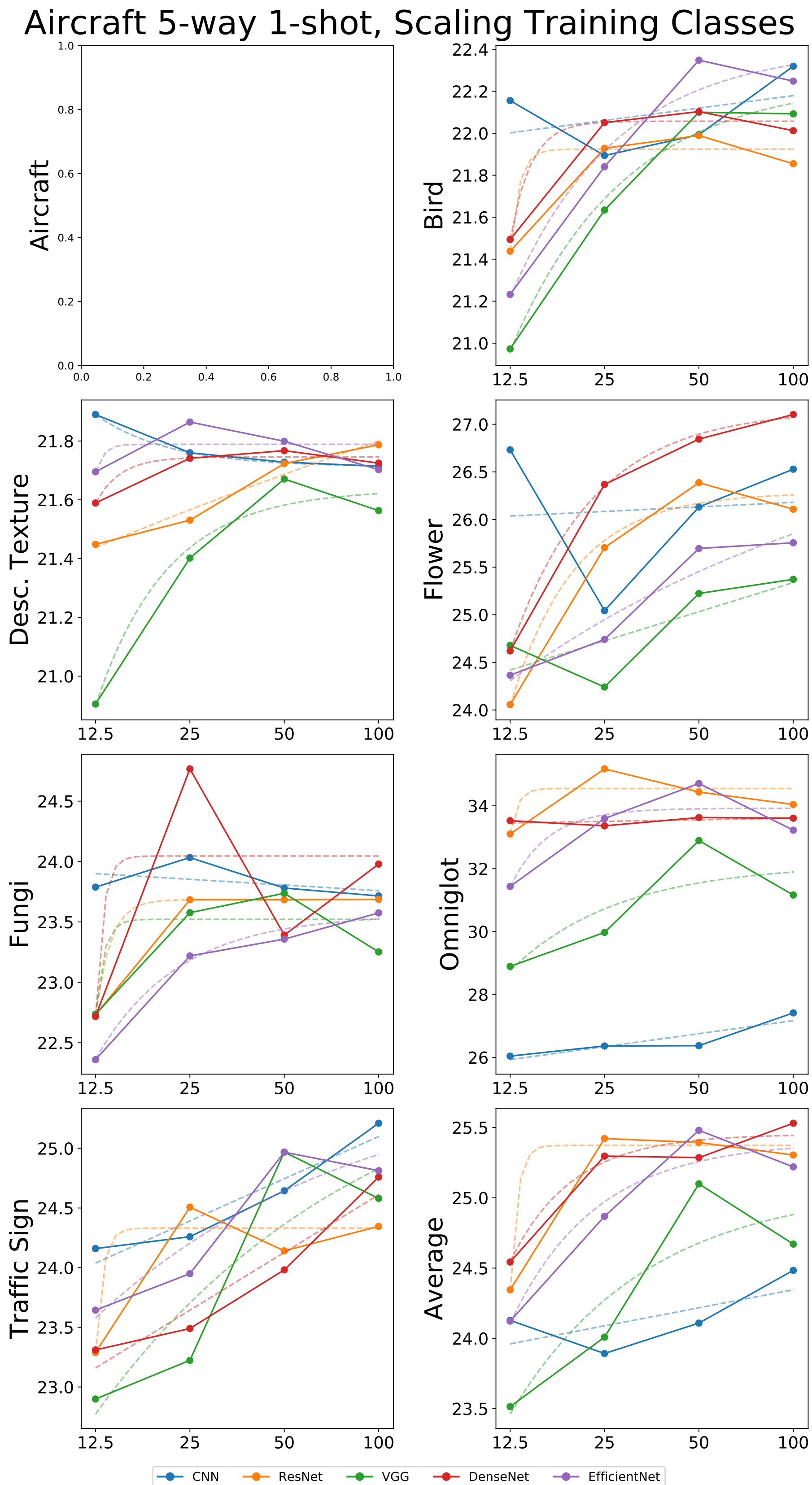

Figure 30. Scaling training classes 5-way 1-shot results for models trained on Aircraft. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 1-shot accuracy.

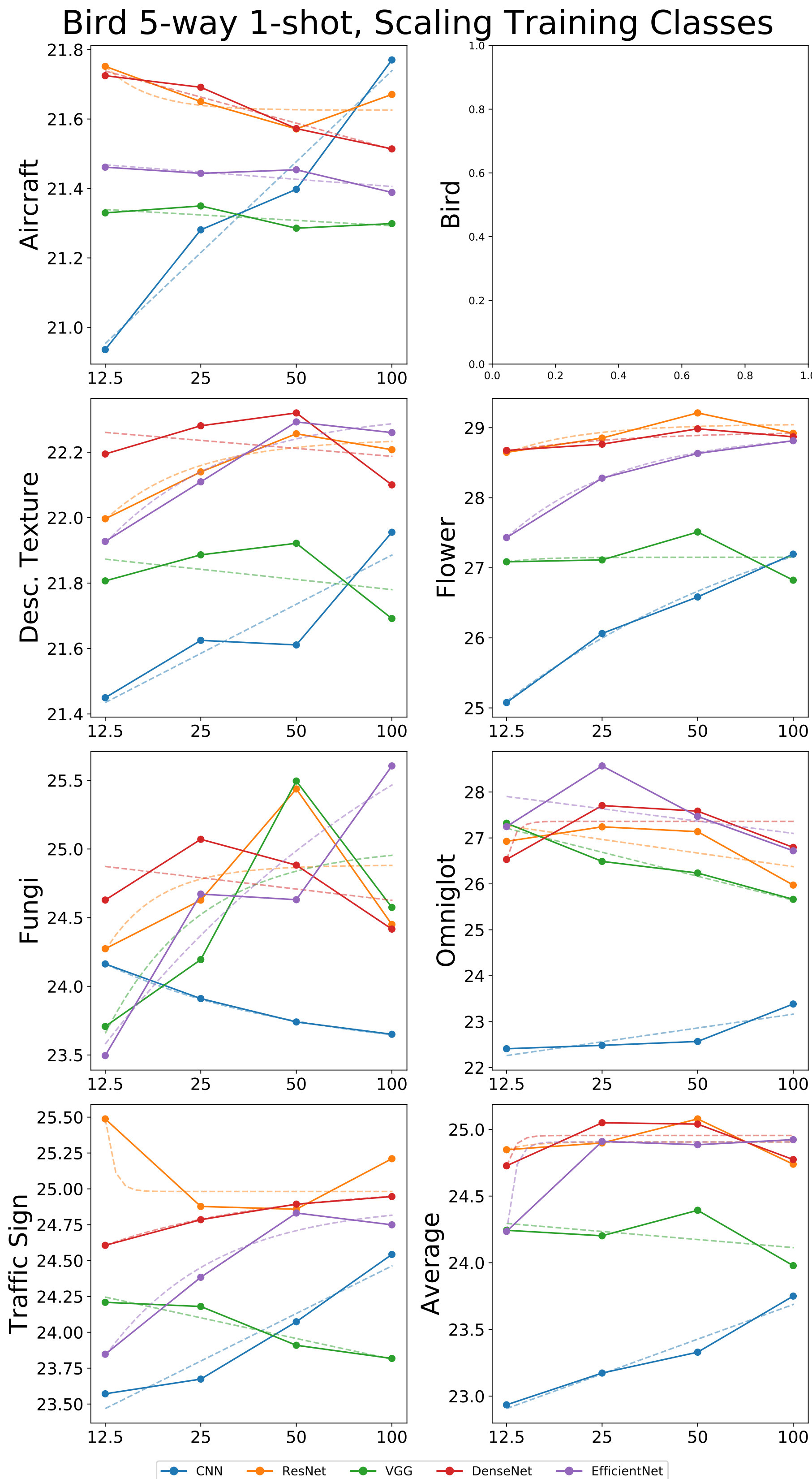

Figure 31. Scaling training classes 5-way 1-shot results for models trained on Bird. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 1-shot accuracy.

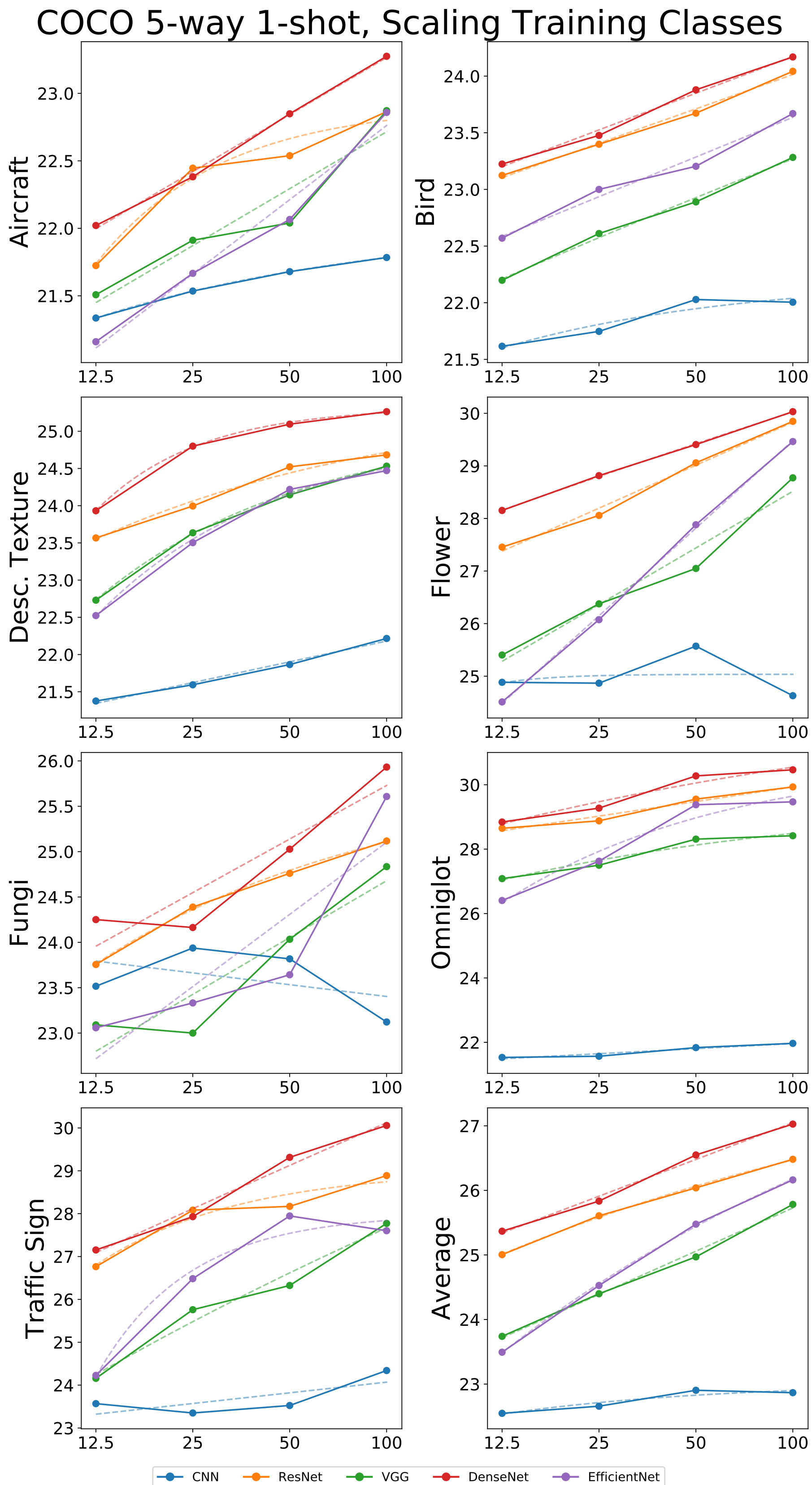

Figure 32. Scaling training classes 5-way 1-shot results for models trained on COCO. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 1-shot accuracy.

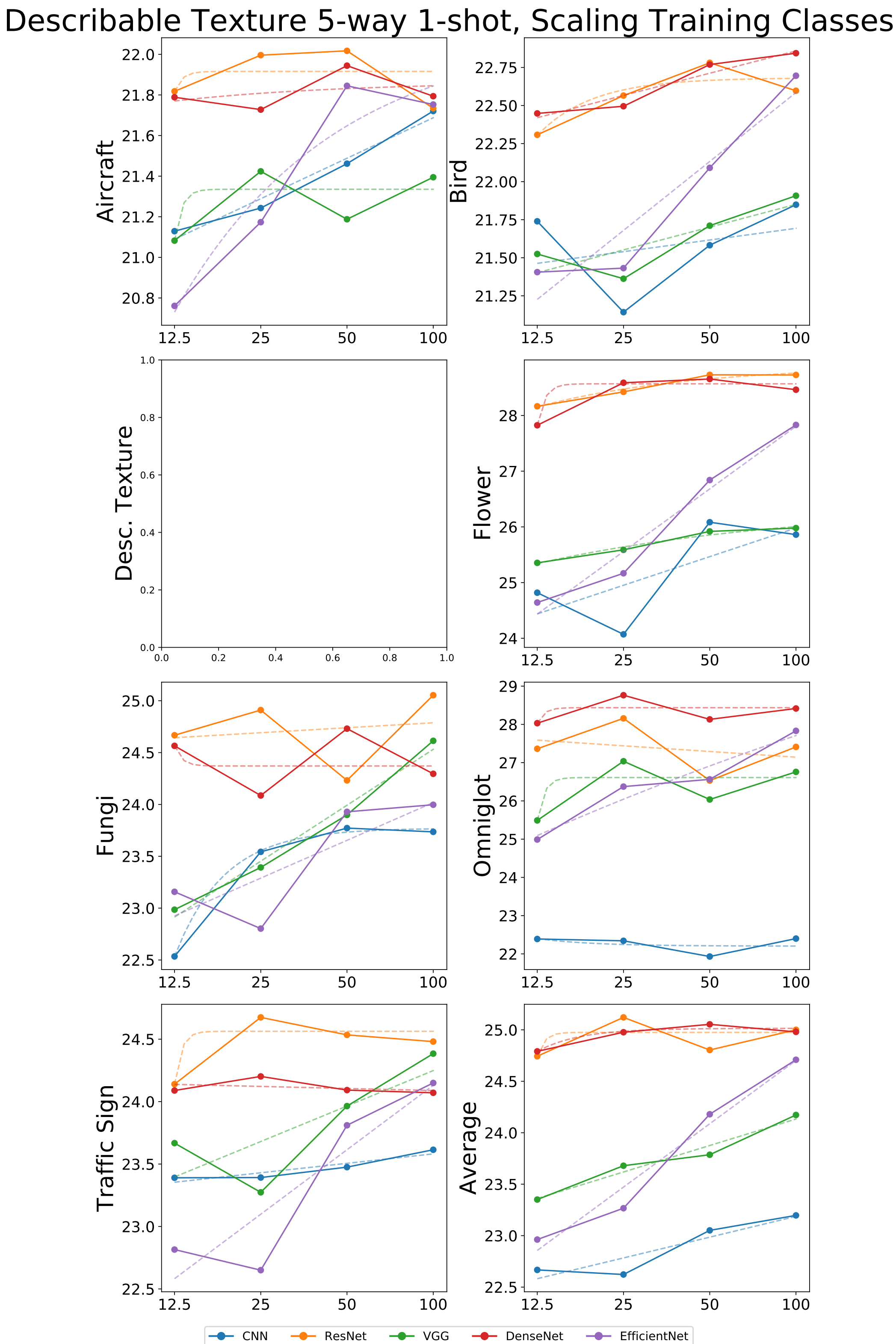

Figure 33. Scaling training classes 5-way 1-shot results for models trained on Describable Texture. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 1-shot accuracy.

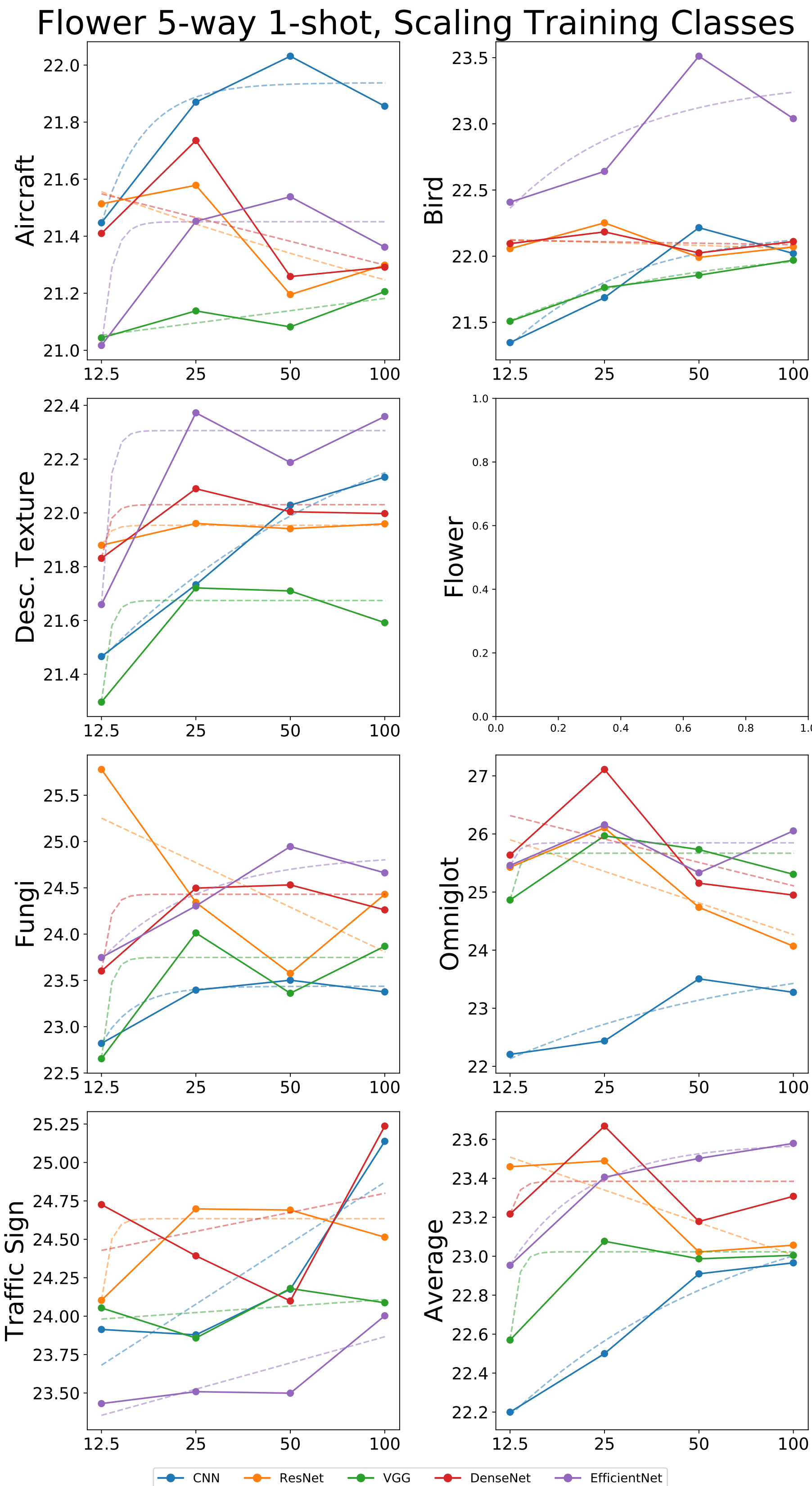

Figure 34. Scaling training classes 5-way 1-shot results for models trained on Flower. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 1-shot accuracy.

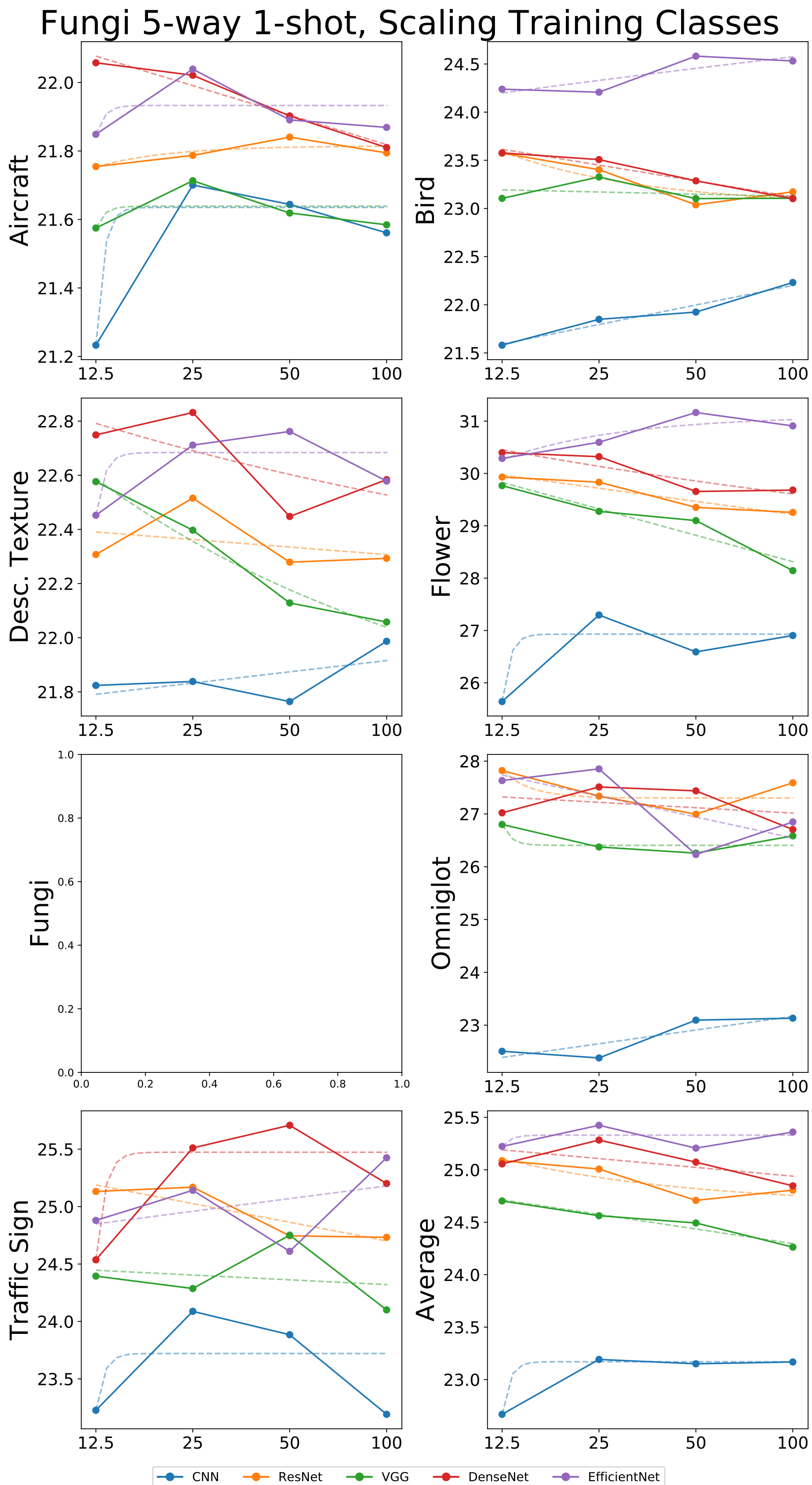

Figure 35. Scaling training classes 5-way 1-shot results for models trained on Fungi. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 1-shot accuracy.

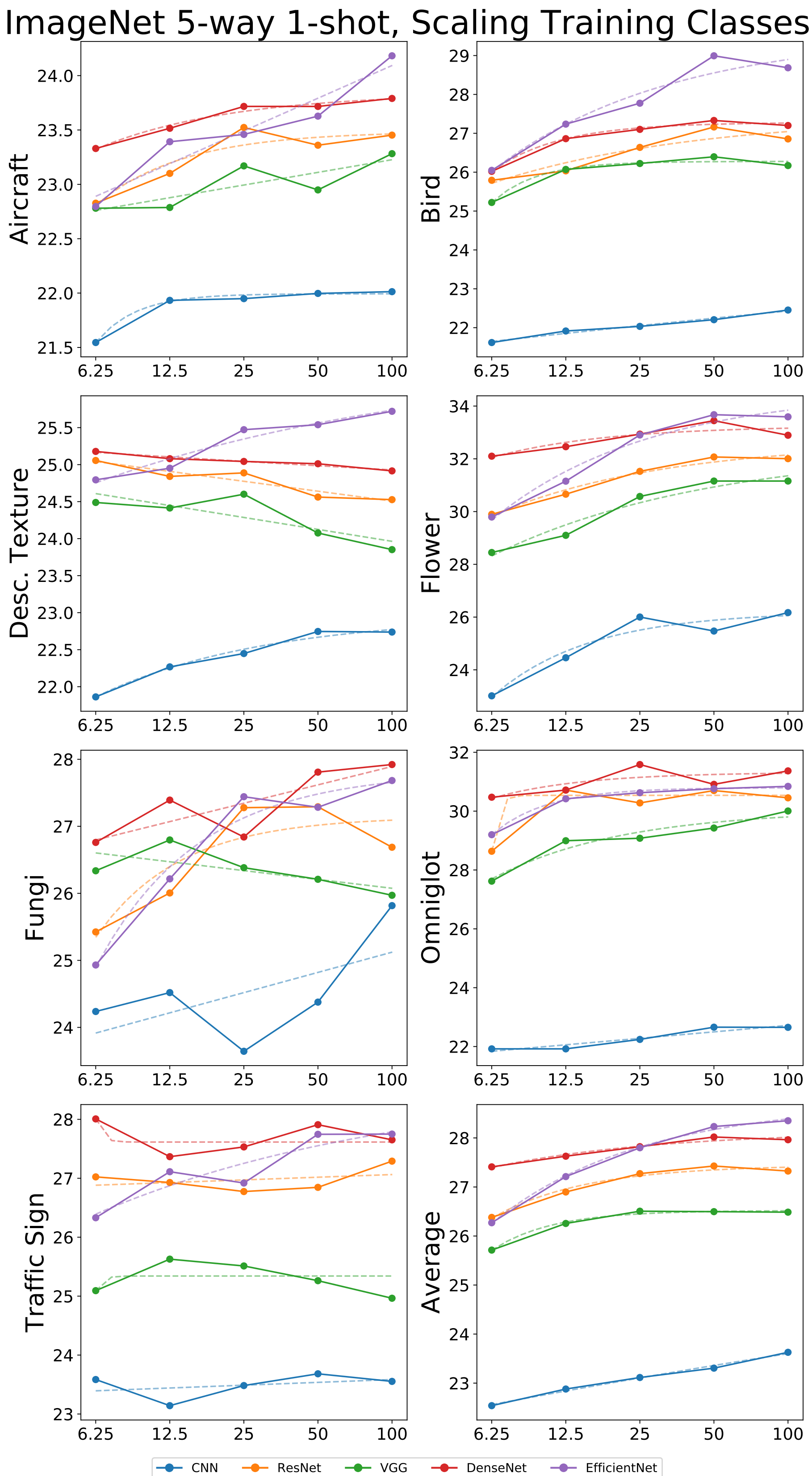

Figure 36. Scaling training classes 5-way 1-shot results for models trained on ImageNet. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 1-shot accuracy.

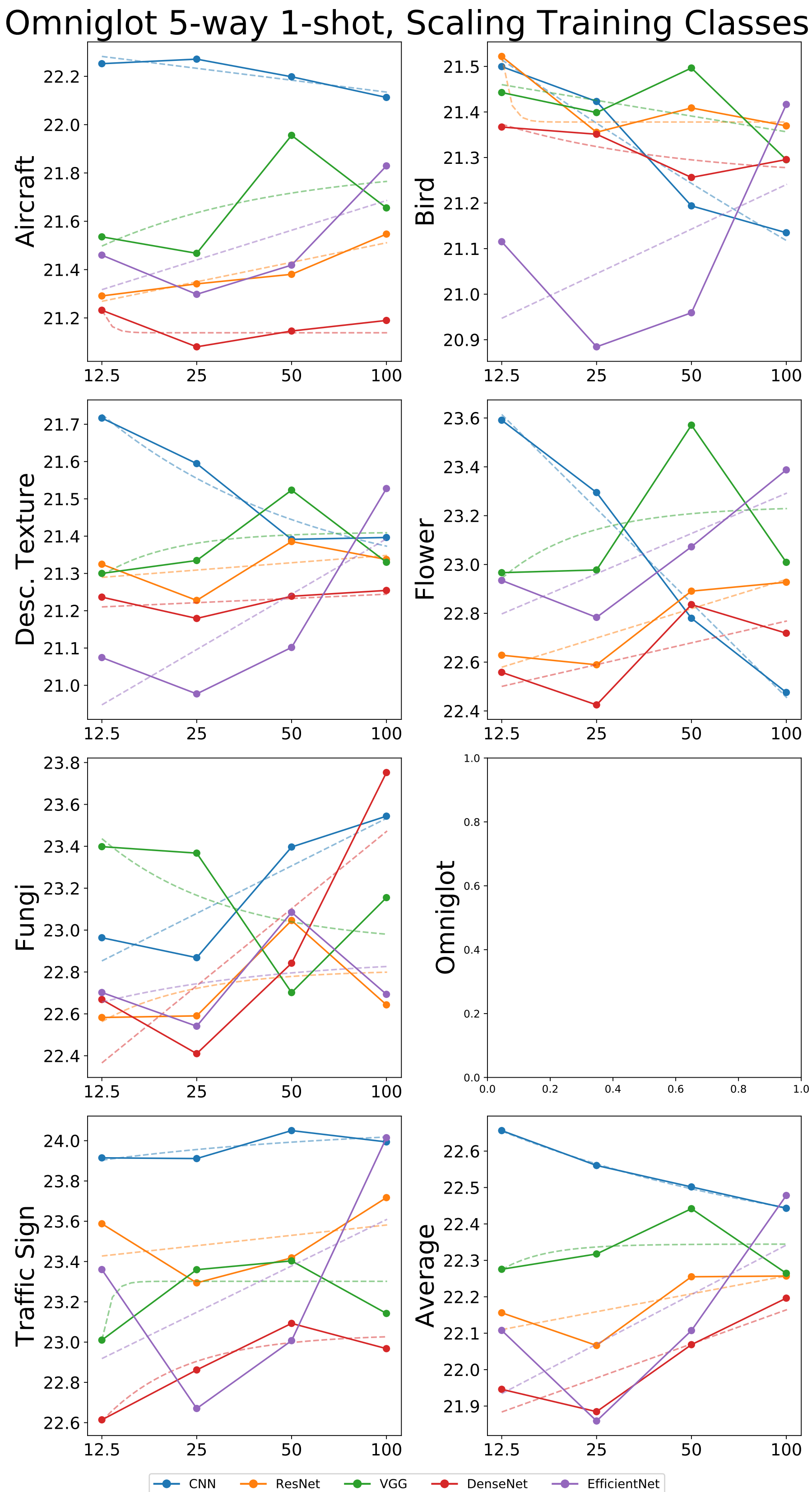

Figure 37. Scaling training classes 5-way 1-shot results for models trained on Omniglot. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 1-shot accuracy.

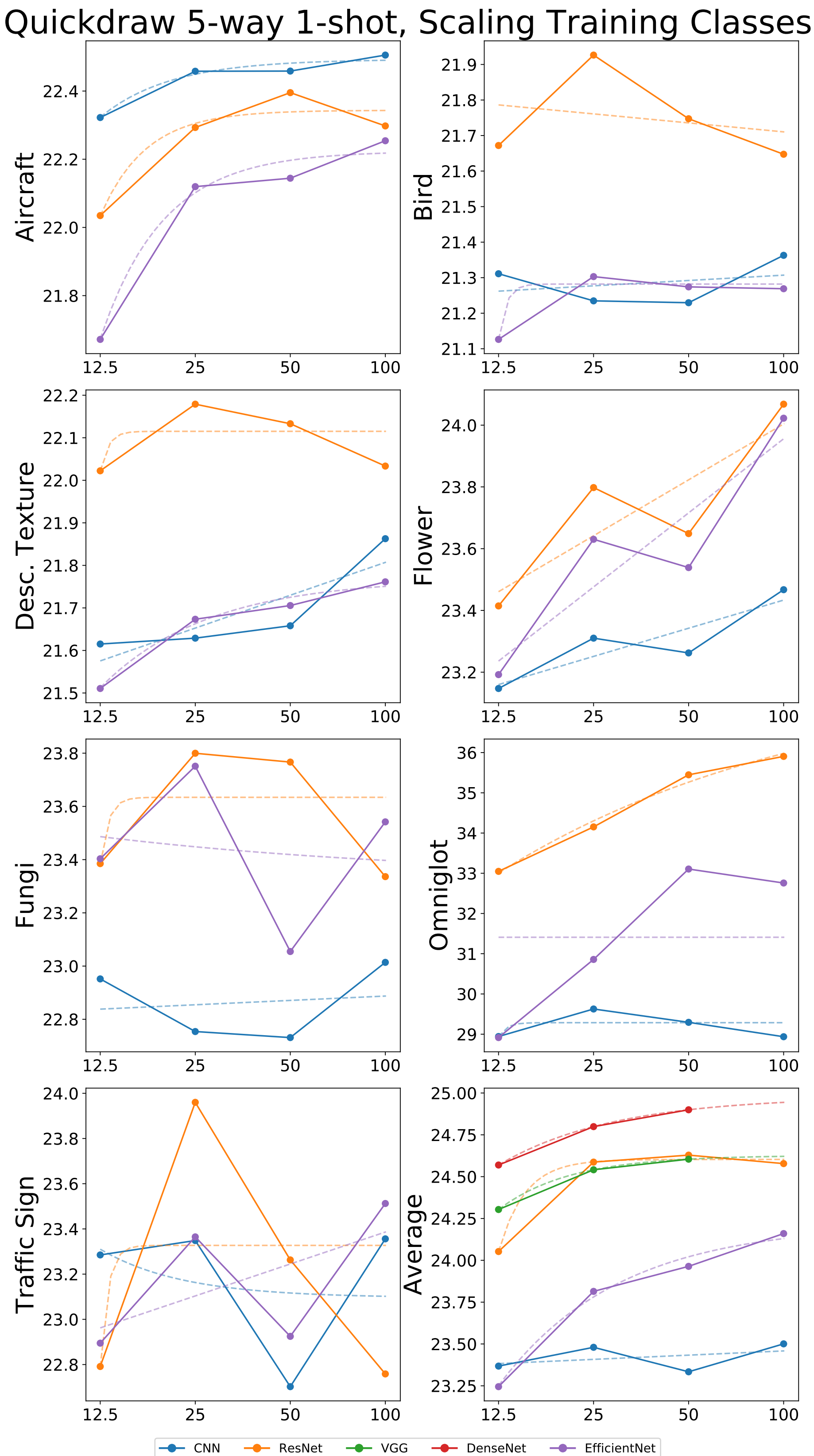

Figure 38. Scaling training classes 5-way 1-shot results for models trained on Quickdraw. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 1-shot accuracy.

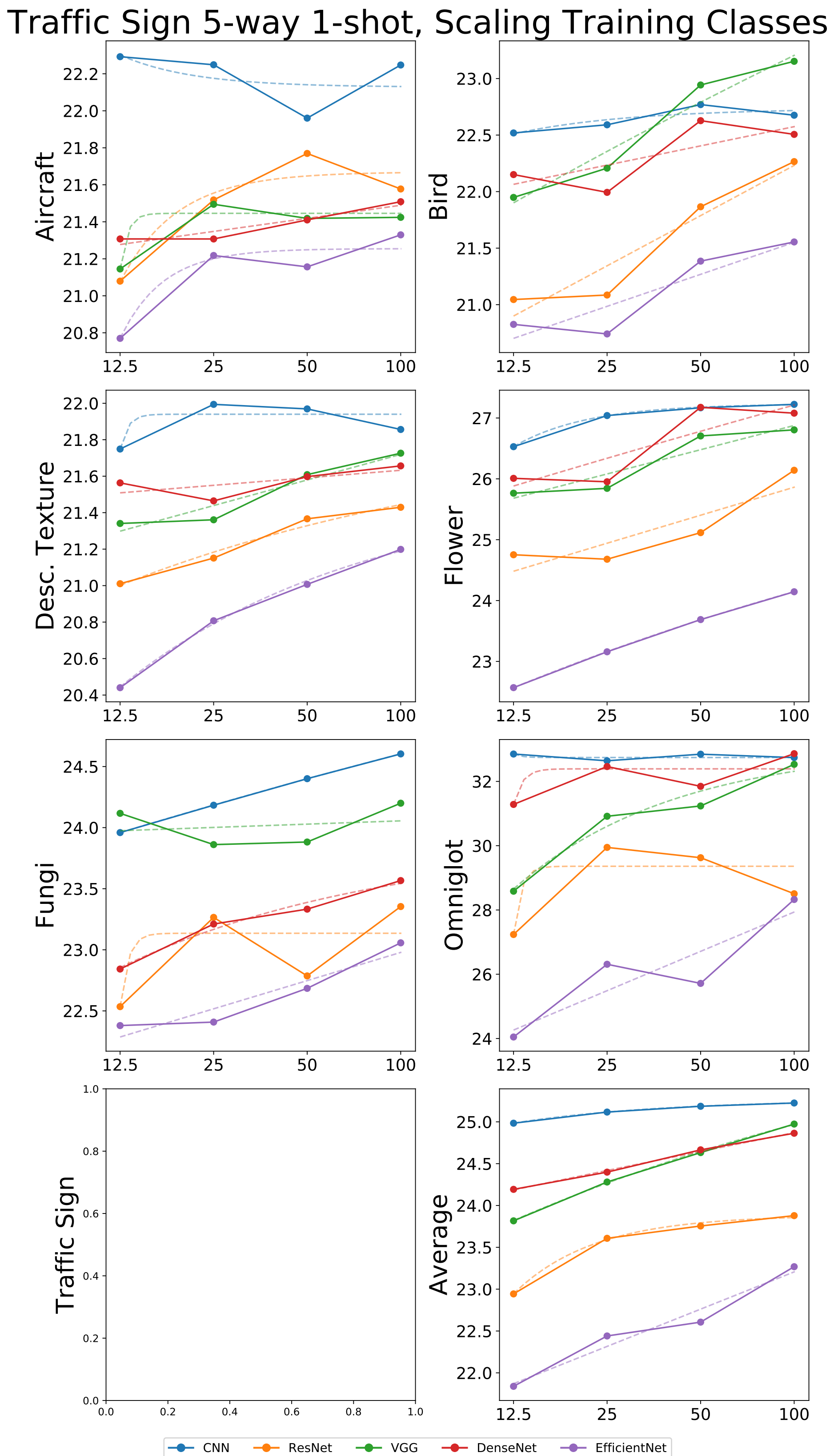

Figure 39. Scaling training classes 5-way 1-shot results for models trained on Traffic Sign. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 1-shot accuracy.

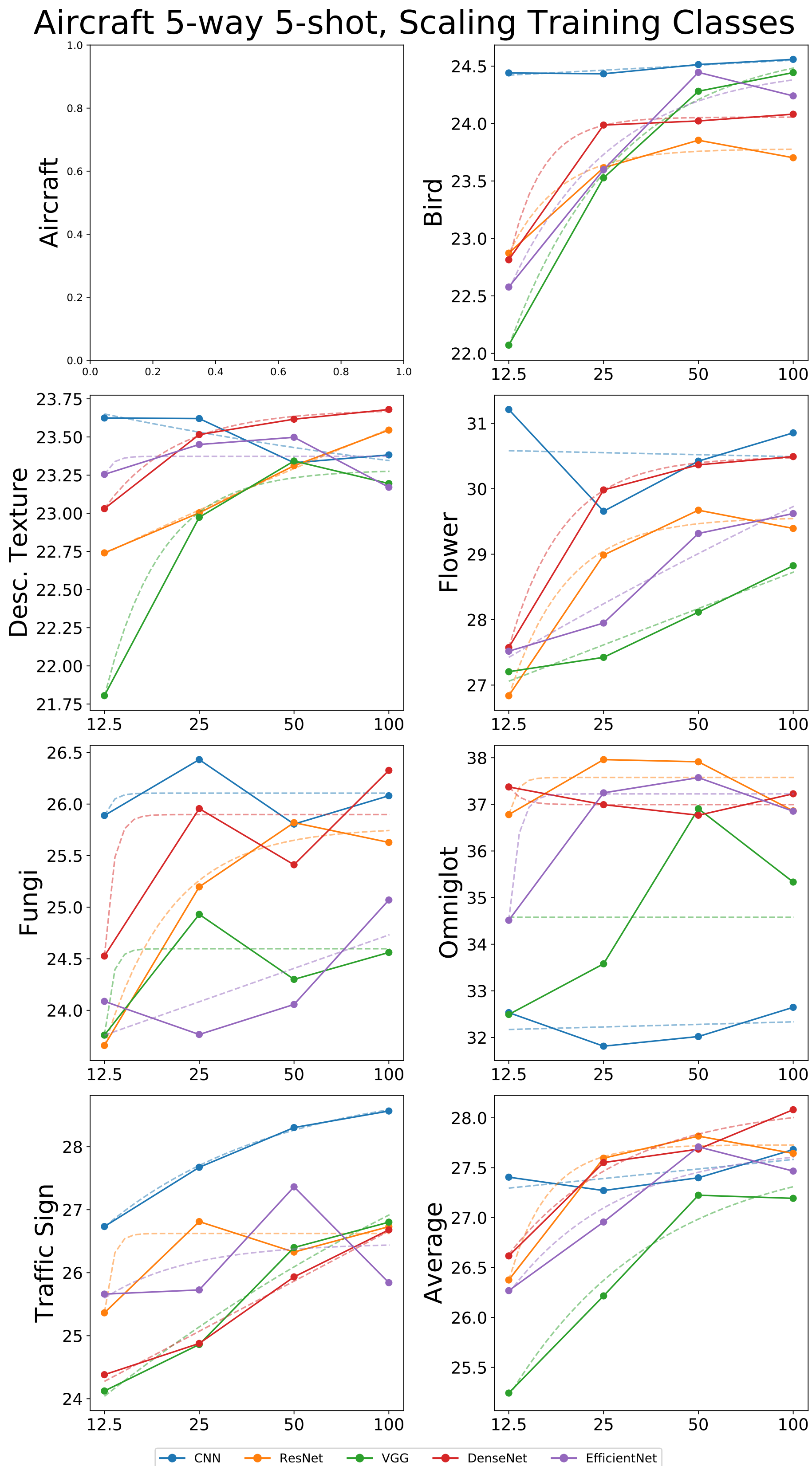

Figure 40. Scaling training classes 5-way 5-shot results for models trained on Aircraft. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 5-shot accuracy.

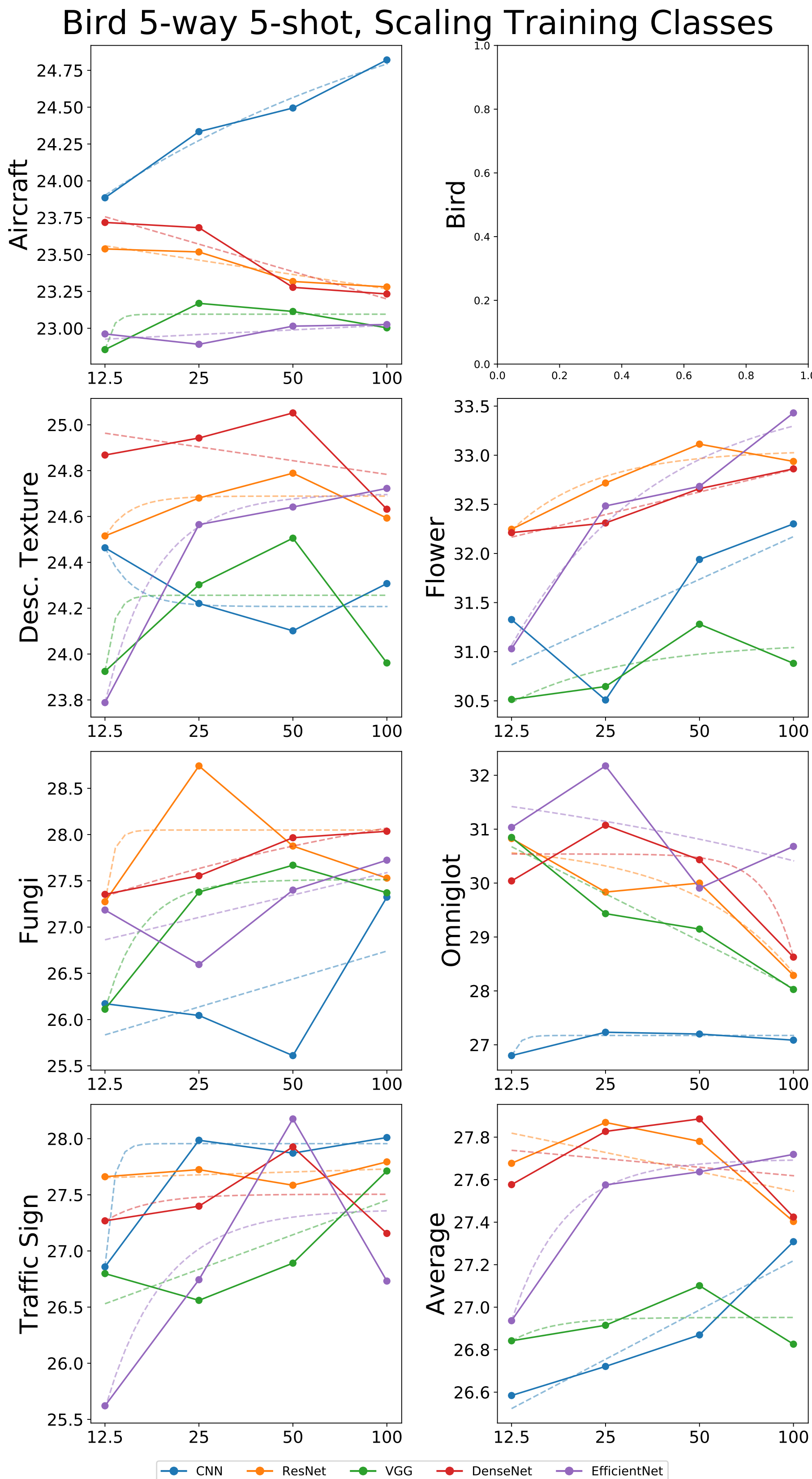

Figure 41. Scaling training classes 5-way 5-shot results for models trained on Bird. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 5-shot accuracy.

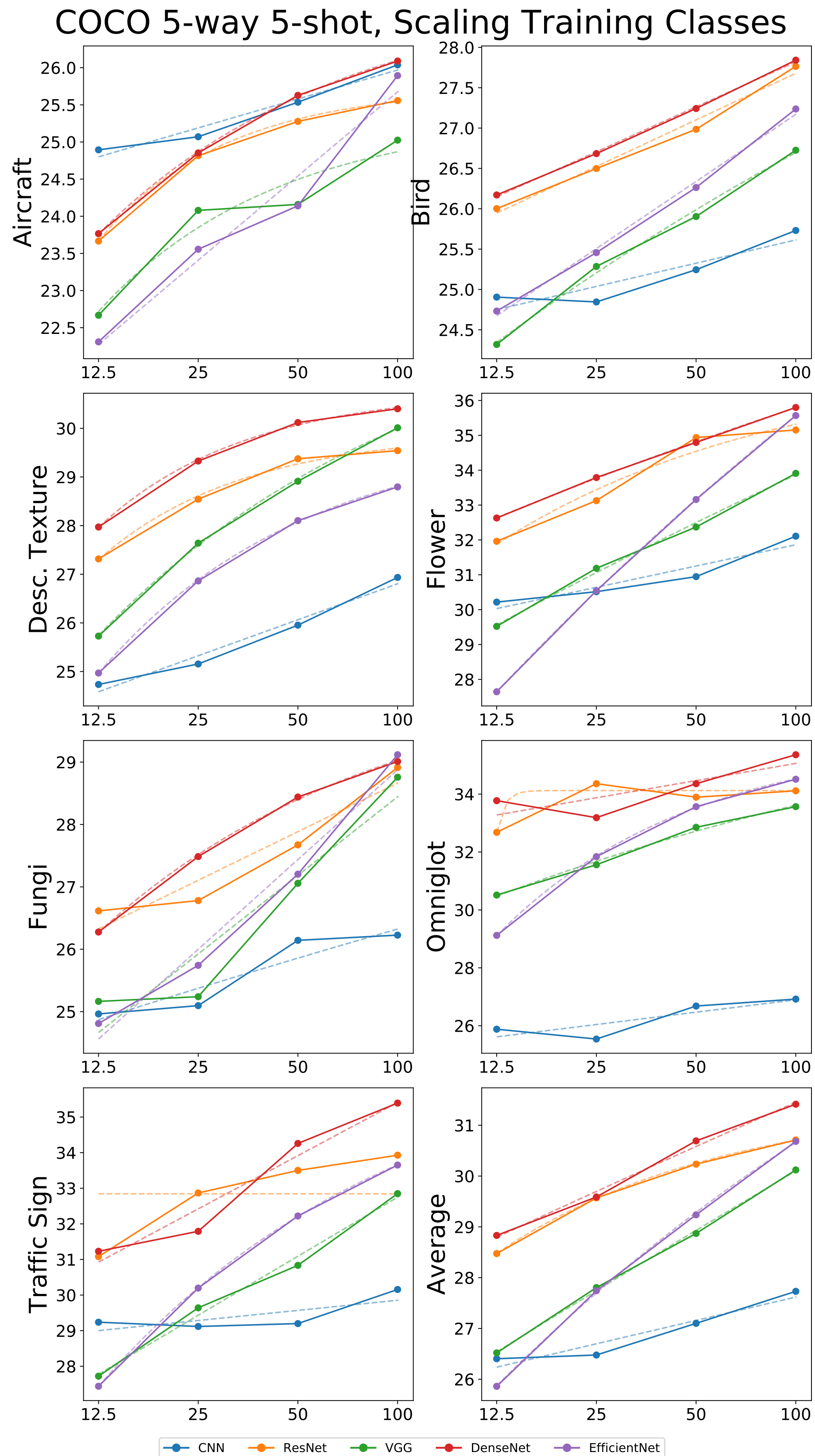

Figure 42. Scaling training classes 5-way 5-shot results for models trained on COCO. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 5-shot accuracy.

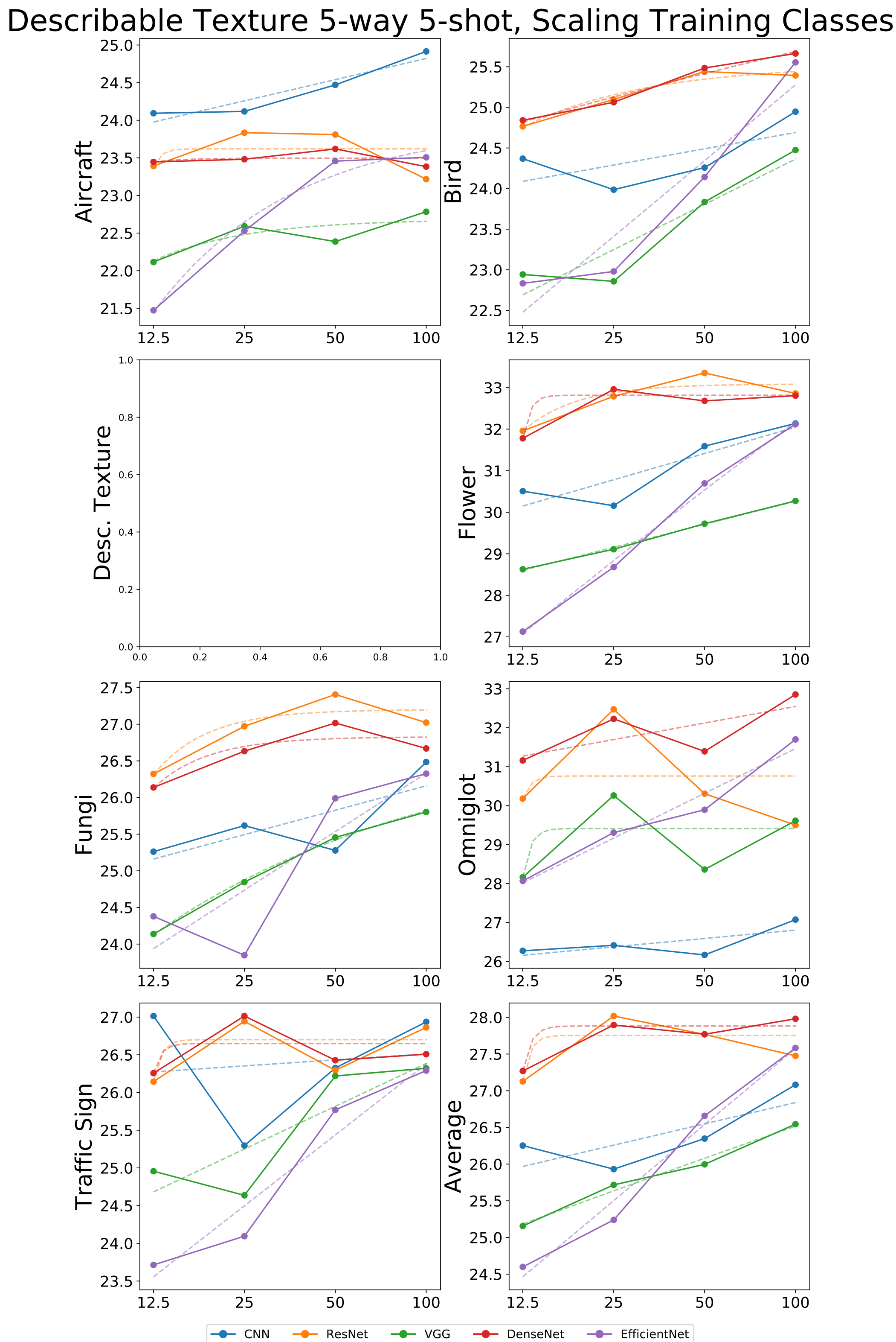

Figure 43. Scaling training classes 5-way 5-shot results for models trained on Describable Texture. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 5-shot accuracy.

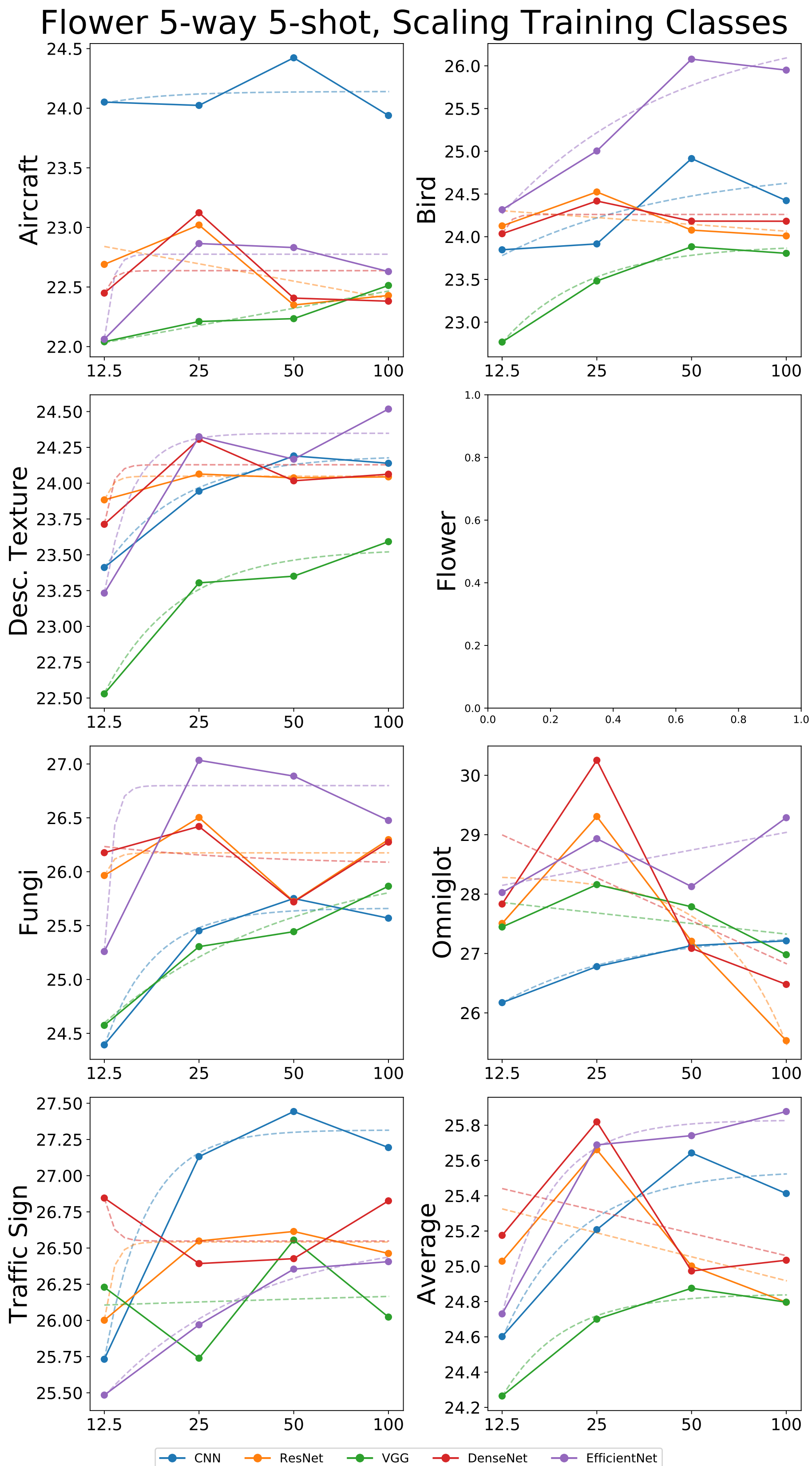

Figure 44. Scaling training classes 5-way 5-shot results for models trained on Flower. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 5-shot accuracy.

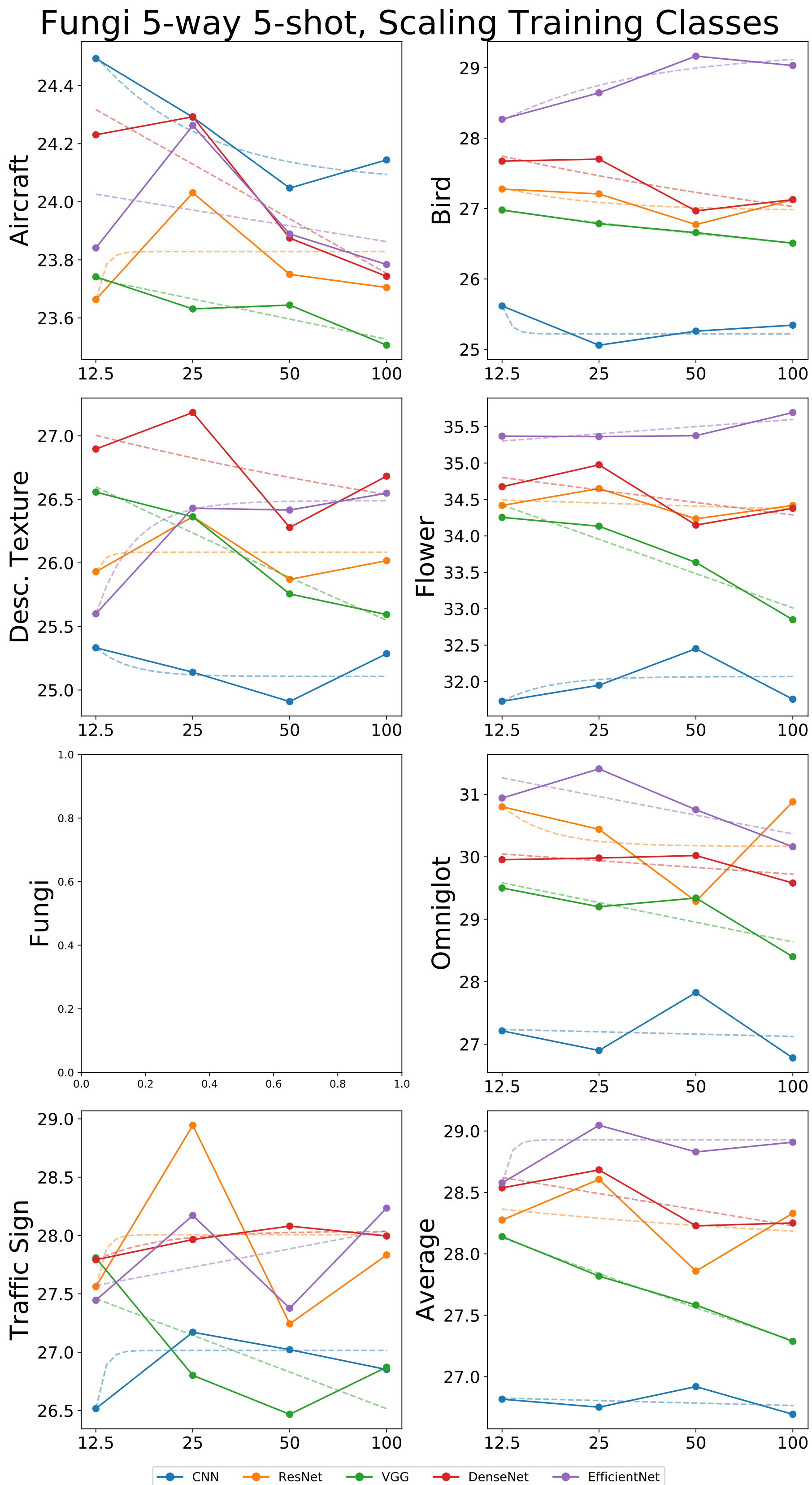

Figure 45. Scaling training classes 5-way 5-shot results for models trained on Fungi. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 5-shot accuracy.

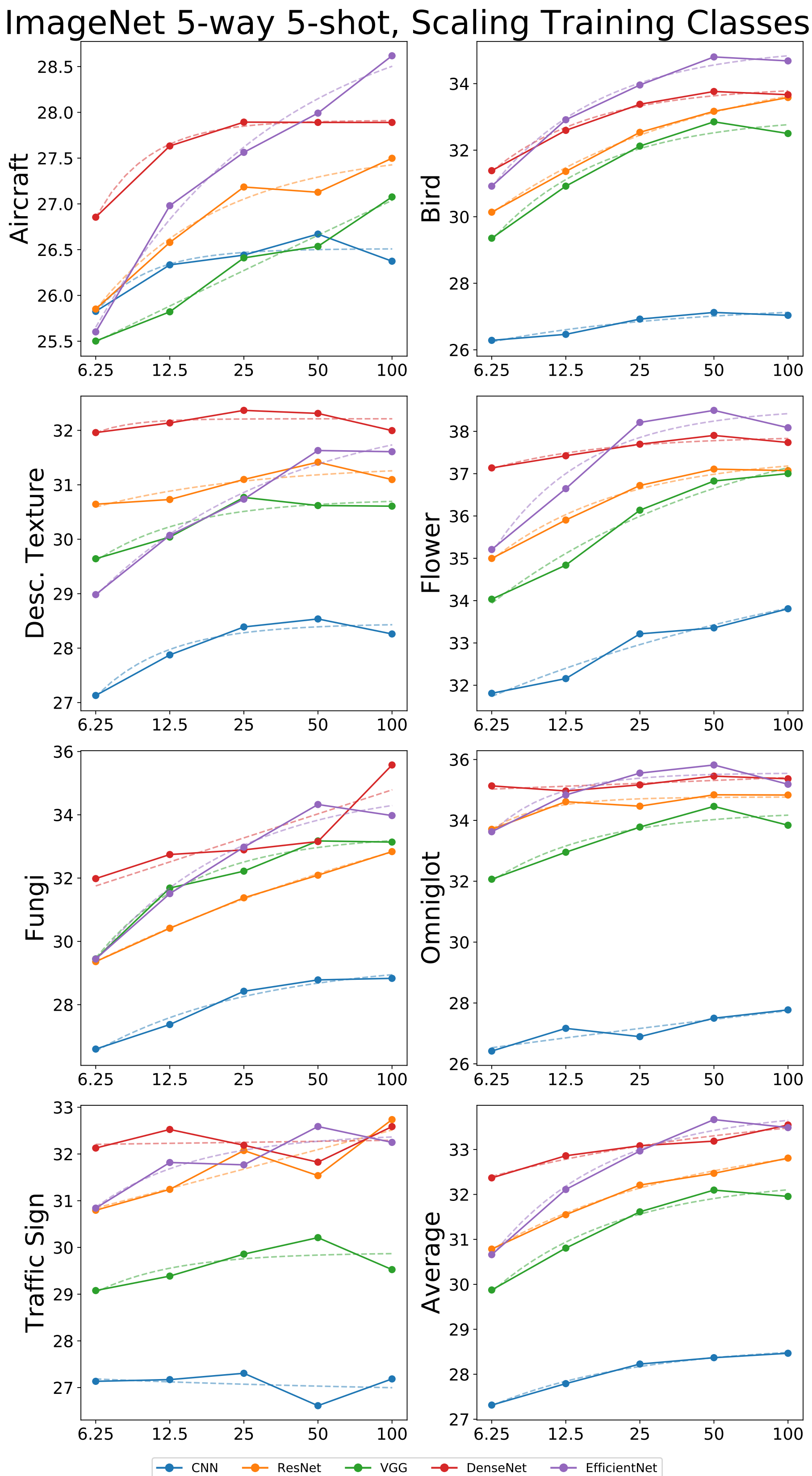

Figure 46. Scaling training classes 5-way 5-shot results for models trained on ImageNet. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 5-shot accuracy.

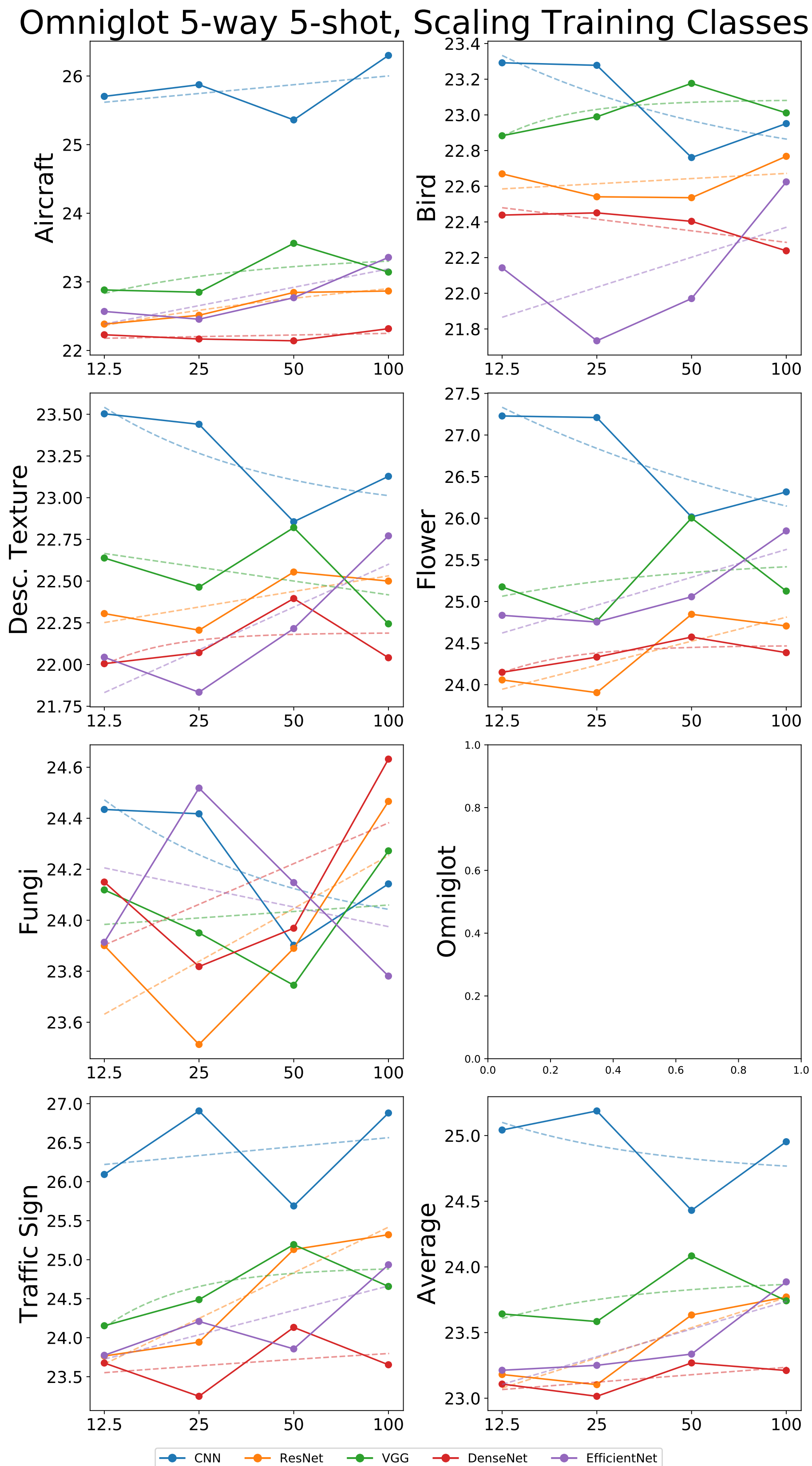

Figure 47. Scaling training classes 5-way 5-shot results for models trained on Omniglot. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 5-shot accuracy.

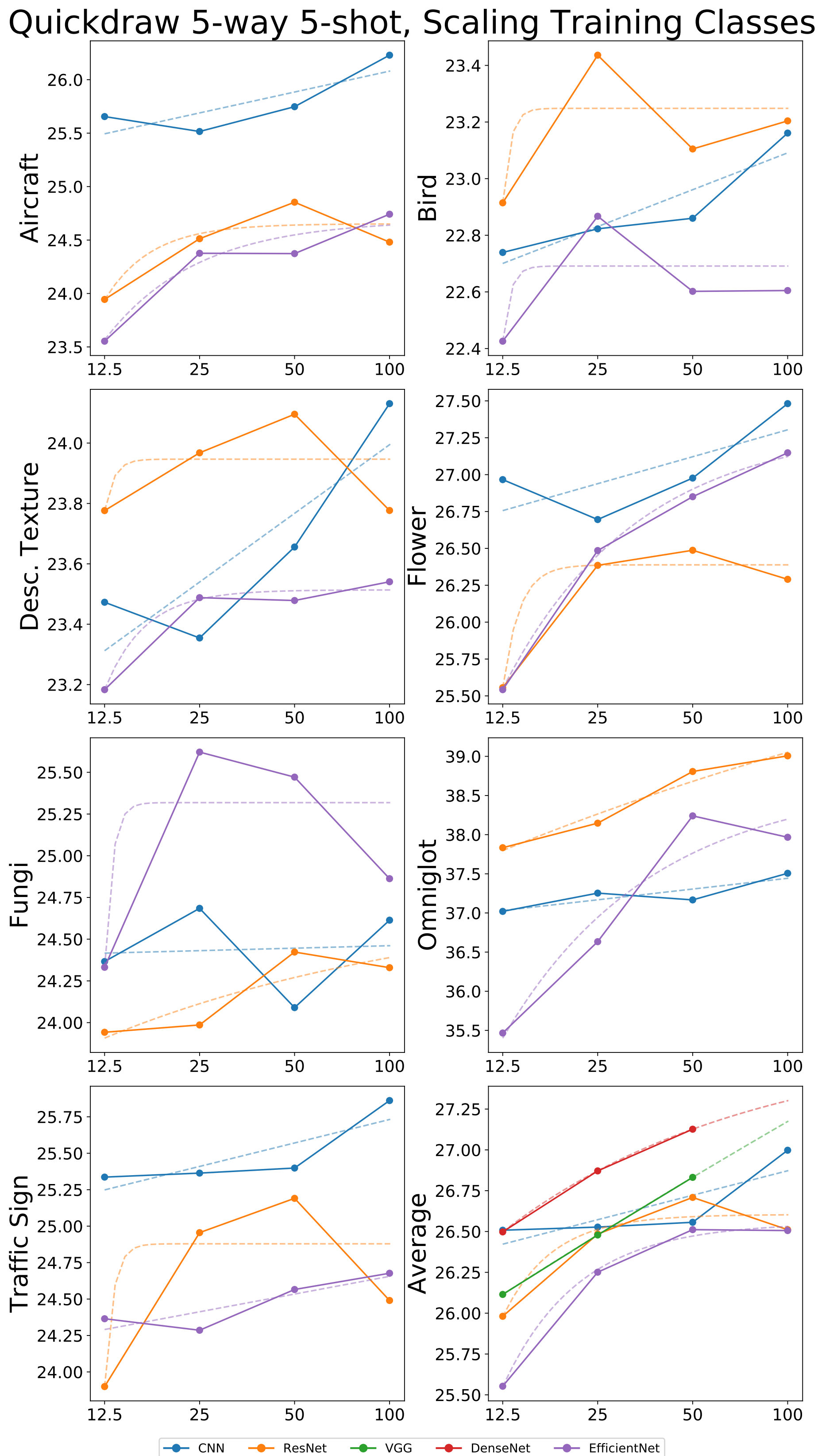

Figure 48. Scaling training classes 5-way 5-shot results for models trained on Quickdraw. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 5-shot accuracy.

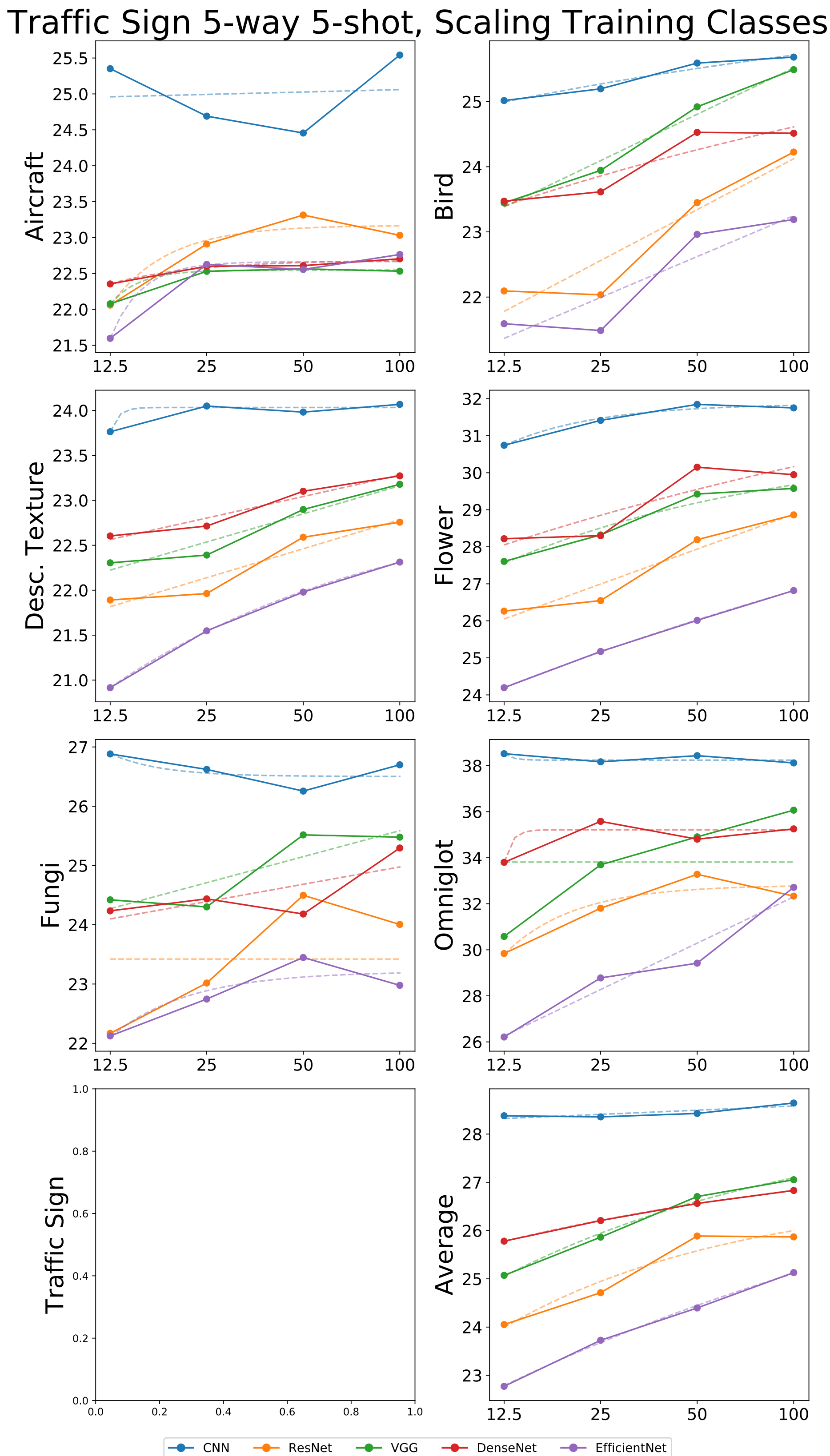

Figure 49. Scaling training classes 5-way 5-shot results for models trained on Traffic Sign. Datasets marked on the left of each plot are the evaluation dataset. Last plot is the average performance. X-axis is the percentage of the total training classes and y-axis is the 5-way 5-shot accuracy.

Table 1. 5-way 1-shot power laws averaged over all 70 train-target pairs.

Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
CNN	$22.1870 + (1.773335)N^{-9.995890\times10^{-2}}$	$1231.7708 + (-1.209191 \times 10^{3})C^{-1.672364\times10^{-4}}$
ResNet	$1592.2211 + (-1.567874 \times 10^{3})N^{-1.263339\times10^{-4}}$	$24.8032 + (-4.427523 \times 10^{3})C^{-3.683409}$
VGG	$2737.7902 + (-2.713921 \times 10^{3})N^{-8.990879\times10^{-5}}$	$24.5747 + (-7.127811)C^{-9.417347\times10^{-1}}$
DenseNet	$2542.9703 + (-2.518490 \times 10^{3})N^{-9.991452\times10^{-5}}$	$25.1753 + (-9.889774)C^{-1.174014}$
EfficientNet	$1939.1736 + (-1.914849 \times 10^{3})N^{-1.089792\times10^{-4}}$	$25.5036 + (-5.956191)C^{-4.918706\times10^{-1}}$

Table 2. 5-way 5-shot power laws averaged over all 70 train-target pairs.

Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
CNN	$27.1226 + (1.344609 \times 10^{13})N^{-1.779765\times10^{1}}$	$2462.0540 + (-2.436143 \times 10^{3})C^{-1.034351\times10^{-4}}$
ResNet	$3227.4282 + (-3.200527 \times 10^{3})N^{-9.933395\times10^{-5}}$	$27.6544 + (-3.130404 \times 10^{2})C^{-2.380884}$
VGG	$4245.4501 + (-4.218864 \times 10^{3})N^{-8.864825\times10^{-5}}$	$27.9630 + (-7.223255)C^{-5.691207\times10^{-1}}$
DenseNet	$4898.5083 + (-4.871348 \times 10^{3})N^{-7.996533\times10^{-5}}$	$28.4222 + (-4.782153)C^{-5.891491\times10^{-1}}$
EfficientNet	$31.1239 + (-4.445441)N^{-1.090730\times10^{-1}}$	$28.5727 + (-1.047196 \times 10^{1})C^{-5.789498\times10^{-1}}$

Table 3. 5-way 1-shot power laws for models trained on Aircraft.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Bird	CNN	$-1929.5322 + (1.952409 \times 10^3)N^{-1.323355\times10^{-4}}$	$960.9739 + (-9.391869 \times 10^2)C^{-9.065506\times10^{-5}}$
	ResNet	$1513.0090 + (-1.491511 \times 10^3)N^{-9.276578\times10^{-5}}$	$21.9243 + (-2.551103 \times 10^{16})C^{-1.524304\times10^{1}}$
	VGG	$22.0547 + (-2.102831 \times 10^{13})N^{-1.706084\times10^{1}}$	$22.2508 + (-2.636380 \times 10^1)C^{-1.195290}$
	DenseNet	$1130.3328 + (-1.108853 \times 10^3)N^{-1.414619\times10^{-4}}$	$22.0574 + (-1.179978 \times 10^7)C^{-6.674216}$
	EfficientNet	$20.7416 + (2.472030)N^{-1.759089\times10^{-1}}$	$22.4101 + (-2.981800 \times 10^1)C^{-1.274974}$
Desc. Text.	CNN	$1043.6049 + (-1.021986 \times 10^3)N^{-7.021320\times10^{-5}}$	$21.7120 + (1.941067 \times 10^1)C^{-1.857344}$
	ResNet	$2662.7046 + (-2.641498 \times 10^3)N^{-7.588005\times10^{-5}}$	$23.5273 + (-2.634425)C^{-9.167206\times10^{-2}}$
	VGG	$2752.3164 + (-2.731441 \times 10^3)N^{-9.422853\times10^{-5}}$	$21.6355 + (-8.672520 \times 10^1)C^{-1.888950}$
	DenseNet	$1112.5782 + (-1.091184 \times 10^3)N^{-9.927059\times10^{-5}}$	$21.7457 + (-1.884023 \times 10^5)C^{-5.543276}$
	EfficientNet	$2576.1039 + (-2.555034 \times 10^3)N^{-1.146341\times10^{-4}}$	$21.7888 + (-9.958388 \times 10^{17})C^{-1.734659\times10^{1}}$
Flower	CNN	$-1738.4170 + (1.765839 \times 10^3)N^{-1.894020\times10^{-4}}$	$594.6899 + (-5.688274 \times 10^2)C^{-1.213515\times10^{-4}}$
	ResNet	$1903.3713 + (-1.877894 \times 10^3)N^{-1.579335\times10^{-4}}$	$26.2837 + (-5.071544 \times 10^2)C^{-2.147864}$
	VGG	$25.5310 + (-3.004669 \times 10^{13})N^{-1.732393\times10^{1}}$	$2364.7425 + (-2.341439 \times 10^3)C^{-1.887628\times10^{-4}}$
	DenseNet	$2122.2164 + (-2.094890 \times 10^3)N^{1.146868\times10^{-4}}$	$27.1539 + (-1.614647 \times 10^2)C^{-1.645475}$
	EfficientNet	$2192.9479 + (-2.166369 \times 10^3)N^{1.073793\times10^{-4}}$	$27.3158 + (-7.210077)C^{-3.459639\times10^{-1}}$
Fungi	CNN	$21.4667 + (3.261498)N^{-1.265862\times10^{-1}}$	$-475.2908 + (4.993625 \times 10^2)C^{-1.361198\times10^{-4}}$
	ResNet	$1321.3129 + (-1.298139 \times 10^3)N^{-1.012673\times10^{-4}}$	$23.6854 + (-1.232363 \times 10^{10})C^{-9.217595}$
	VGG	$-1045.2132 + (1.069096 \times 10^3)N^{-1.204776\times10^{-4}}$	$23.5219 + (-7.680866 \times 10^{17})C^{-1.640058\times10^{1}}$
	DenseNet	$-657.7067 + (6.820683 \times 10^2)N^{-1.372123\times10^{-4}}$	$24.0459 + (-1.301472 \times 10^{20})C^{-1.822504\times10^{1}}$
	EfficientNet	$-237.1656 + (2.611879 \times 10^2)N^{-4.208639\times10^{-4}}$	$23.5626 + (-7.739143 \times 10^1)C^{-1.650515}$
Omniglot	CNN	$1568.9569 + (-1.541200 \times 10^3)N^{1.432579\times10^{-4}}$	$4126.0904 + (-4.101667 \times 10^3)C^{-1.453926\times10^{-4}}$
	ResNet	$35.0053 + (-1.422807 \times 10^{15})N^{-1.874159\times10^{1}}$	$34.5509 + (-2.633291 \times 10^{20})C^{-1.847048\times10^{1}}$
	VGG	$36.9287 + (-7.526573)N^{-1.504027\times10^{-1}}$	$32.1485 + (-7.857978 \times 10^1)C^{-1.245213}$
	DenseNet	$39.0521 + (-1.240233 \times 10^1)N^{-2.567592\times10^{-1}}$	$369.9616 + (-3.366886 \times 10^2)C^{-2.143989\times10^{-4}}$
	EfficientNet	$34.1255 + (-3.836734 \times 10^{-4})N^{1.634717}$	$33.9203 + (-2.176470 \times 10^4)C^{-3.593608}$
Traffic Sign	CNN	$25.9134 + (-3.568267 \times 10^{-1})N^{2.666234\times10^{-1}}$	$3965.0534 + (-3.942304 \times 10^3)C^{-1.294695\times10^{-4}}$
	ResNet	$771.8651 + (-7.479907 \times 10^2)N^{-1.055483\times10^{-4}}$	$24.3314 + (-4.781904 \times 10^{20})C^{-1.883609\times10^{1}}$
	VGG	$24.2355 + (-6.465501 \times 10^{14})N^{-1.870357\times10^{1}}$	$25.9656 + (-1.120540 \times 10^1)C^{-4.971646\times10^{-1}}$
	DenseNet	$24.5496 + (-5.884418)N^{-1.511095}$	$7832.4032 + (-7.811007 \times 10^3)C^{-8.938921\times10^{-5}}$
	EfficientNet	$403.6154 + (-3.801087 \times 10^2)N^{-2.682043\times10^{-4}}$	$25.6547 + (-7.695378)C^{-5.185766\times10^{-1}}$
Average	CNN	$23.9211 + (2.024914)N^{-5.386192\times10^{-1}}$	$1542.0394 + (-1.518547 \times 10^3)C^{-1.220469\times10^{-4}}$
	ResNet	$1106.6910 + (-1.081852 \times 10^3)N^{-1.714032\times10^{-4}}$	$25.3731 + (-1.417491 \times 10^{21})C^{-1.927178\times10^{1}}$
	VGG	$24.9340 + (-1.264069 \times 10^3)N^{-4.186695}$	$25.0848 + (-2.018792 \times 10^1)C^{-9.980007\times10^{-1}}$
	DenseNet	$1280.3629 + (-1.255769 \times 10^3)N^{-1.924645\times10^{-4}}$	$25.4537 + (-2.273473 \times 10^2)C^{-2.187322}$
	EfficientNet	$25.0530 + (5.450569 \times 10^{-1})N^{-3.541147\times10^{-1}}$	$25.4007 + (-7.187890 \times 10^1)C^{-1.590799}$

Table 4. 5-way 1-shot power laws for models trained on Bird.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$-1775.4777 + (1.797718 \times 10^3)N^{-1.228546\times10^{-4}}$	$1150.2426 + (-1.130245 \times 10^3)C^{-3.347491\times10^{-4}}$
	ResNet	$1057.8324 + (-1.036680 \times 10^3)N^{-1.931767\times10^{-4}}$	$21.6253 + (3.740832 \times 10^2)C^{-3.162532}$
	VGG	$2082.1145 + (-2.061393 \times 10^3)N^{-1.031245\times10^{-4}}$	$-25.0257 + (4.642239 \times 10^1)C^{-4.880303\times10^{-4}}$
	DenseNet	$3193.2815 + (-3.172487 \times 10^3)N^{-9.797140\times10^{-5}}$	$-399.5662 + (4.215787 \times 10^2)C^{-2.574427\times10^{-4}}$
	EfficientNet	$1384.8663 + (-1.363615 \times 10^3)N^{-9.613991\times10^{-5}}$	$-133.5660 + (1.551097 \times 10^2)C^{-1.933779\times10^{-4}}$
Desc. Text.	CNN	$70.5487 + (-4.866876 \times 10^1)N^{-3.418563\times10^{-4}}$	$1263.5050 + (-1.242618 \times 10^3)C^{-1.746255\times10^{-4}}$
	ResNet	$22.3801 + (-2.080148)N^{-1.043867}$	$22.2420 + (-1.364947 \times 10^1)C^{-1.586452}$
	VGG	$3402.2021 + (-3.381157 \times 10^3)N^{-8.680383\times10^{-5}}$	$-401.8851 + (4.238712 \times 10^2)C^{-1.055583\times10^{-4}}$
	DenseNet	$2842.5832 + (-2.821044 \times 10^3)N^{-8.521555\times10^{-5}}$	$-271.9371 + (2.942865 \times 10^2)C^{-1.195903\times10^{-4}}$
	EfficientNet	$23.4673 + (-1.893415)N^{-1.826938\times10^{-1}}$	$22.3264 + (-6.846331)C^{-1.119681}$
Flower	CNN	$14.1680 + (1.584145 \times 10^1)N^{-7.170788\times10^{-2}}$	$28.5944 + (-1.036558 \times 10^1)C^{-4.298389\times10^{-1}}$
	ResNet	$29.9234 + (-5.263883)N^{-6.255683\times10^{-1}}$	$29.0527 + (-3.826956 \times 10^1)C^{-1.791838}$
	VGG	$2473.8000 + (-2.446851 \times 10^3)N^{-8.872708\times10^{-5}}$	$27.1506 + (-3.203672 \times 10^3)C^{-4.276271}$
	DenseNet	$3422.1965 + (-3.394854 \times 10^3)N^{-1.827288\times10^{-4}}$	$28.9461 + (-5.161771)C^{-1.153189}$
	EfficientNet	$30.8928 + (-4.635177)N^{-3.534832\times10^{-1}}$	$28.9376 + (-2.961904 \times 10^1)C^{-1.179740}$
Fungi	CNN	$-716.3445 + (7.403644 \times 10^2)N^{-1.709149\times10^{-4}}$	$23.4880 + (3.892102)C^{-6.929850\times10^{-1}}$
	ResNet	$24.9418 + (2.370727 \times 10^{12})N^{-1.648266\times10^{1}}$	$24.8831 + (-4.453670 \times 10^2)C^{-2.603307}$
	VGG	$5446.6576 + (-5.424207 \times 10^3)N^{-1.077717\times10^{-4}}$	$25.0205 + (-5.348049 \times 10^1)C^{-1.452797}$
	DenseNet	$813.0244 + (-7.898304 \times 10^2)N^{-7.020523\times10^{-4}}$	$-920.1177 + (9.452905 \times 10^2)C^{-1.255862\times10^{-4}}$
	EfficientNet	$2433.7248 + (-2.409245 \times 10^3)N^{-9.322441\times10^{-5}}$	$27.1513 + (-8.891914)C^{-3.611532\times10^{-1}}$
Omniglot	CNN	$22.2715 + (2.866387)N^{-4.036425\times10^{-1}}$	$3179.4403 + (-3.158275 \times 10^3)C^{-1.373389\times10^{-4}}$
	ResNet	$7352.4724 + (-7.328801 \times 10^3)N^{-1.628489\times10^{-4}}$	$-2335.1546 + (2.363499 \times 10^3)C^{-1.809686\times10^{-4}}$
	VGG	$1840.1742 + (-1.816324 \times 10^3)N^{-4.492081\times10^{-4}}$	$2964.7232 + (-2.935608 \times 10^3)C^{2.566061\times10^{-4}}$
	DenseNet	$9566.0237 + (-9.542314 \times 10^3)N^{-1.142279\times10^{-4}}$	$27.3614 + (-1.615931 \times 10^{19})C^{-1.758526\times10^{1}}$
	EfficientNet	$8014.2028 + (-7.988932 \times 10^3)N^{-9.925759\times10^{-5}}$	$-2959.2721 + (2.988151 \times 10^3)C^{-1.293759\times10^{-4}}$
Traffic Sign	CNN	$23.9746 + (7.754703)N^{-1.163037}$	$3850.0653 + (-3.827805 \times 10^3)C^{-1.249694\times10^{-4}}$
	ResNet	$25.1386 + (-1.401922 \times 10^2)N^{-2.867489}$	$24.9818 + (7.947418 \times 10^{18})C^{-1.750029\times10^{1}}$
	VGG	$24.4150 + (-1.403413)N^{-4.908663\times10^{-1}}$	$-649.6771 + (6.744494 \times 10^2)C^{-3.092605\times10^{-4}}$
	DenseNet	$25.5485 + (-1.419394)N^{-2.151248\times10^{-1}}$	$25.0232 + (-3.345218)C^{-8.240048\times10^{-1}}$
	EfficientNet	$25.1331 + (-5.163718 \times 10^1)N^{-2.129529}$	$24.8936 + (-2.541642 \times 10^1)C^{-1.258538}$
Average	CNN	$-1311.6345 + (1.336181 \times 10^3)N^{-2.329917\times10^{-4}}$	$2874.5497 + (-2.852592 \times 10^3)C^{-1.316713\times10^{-4}}$
	ResNet	$71.1129 + (-4.719938 \times 10^1)N^{-7.919413\times10^{-3}}$	$24.9056 + (-1.184331 \times 10^5)C^{-5.750935}$
	VGG	$2442.1101 + (-2.419002 \times 10^3)N^{-1.561016\times10^{-4}}$	$-468.8266 + (4.933418 \times 10^2)C^{-1.768068\times10^{-4}}$
	DenseNet	$4800.1510 + (-4.776666 \times 10^3)N^{-1.035922\times10^{-4}}$	$24.9543 + (-2.364241 \times 10^{18})C^{-1.733672\times10^{1}}$
	EfficientNet	$3737.7566 + (-3.713718 \times 10^3)N^{-9.822943\times10^{-5}}$	$24.9059 + (-2.646171 \times 10^{19})C^{-1.786473\times10^{1}}$

Table 5. 5-way 1-shot power laws for models trained on COCO.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$-455.8771 + (4.778011 \times 10^{2})N^{-1.052265\times10^{-4}}$	$22.0560 + (-2.348949)C^{-4.679771\times10^{-1}}$
	ResNet	$22.9997 + (-9.021319)N^{-1.574895}$	$22.9152 + (-1.966180 \times 10^{1})C^{-1.114694}$
	VGG	$628.9665 + (-6.064218 \times 10^{2})N^{-1.426832\times10^{-4}}$	$4462.3090 + (-4.442397 \times 10^{3})C^{-1.370921\times10^{-4}}$
	DenseNet	$23.4106 + (-1.758315 \times 10^{1})N^{-1.769762}$	$4341.3194 + (-4.320863 \times 10^{3})C^{-1.412357\times10^{-4}}$
	EfficientNet	$23.0690 + (-8.472271)N^{-1.281701}$	$4400.8645 + (-4.381754 \times 10^{3})C^{-1.810225\times10^{-4}}$
Bird	CNN	$22.3681 + (-1.431713 \times 10^{-1})N^{1.065097\times10^{-5}}$	$22.2228 + (-2.744657)C^{-5.875699\times10^{-1}}$
	ResNet	$24.3115 + (-2.084742)N^{-5.040176\times10^{-1}}$	$3941.9556 + (-3.919956 \times 10^{3})C^{-1.115911\times10^{-4}}$
	VGG	$23.8224 + (-1.889336)N^{-3.878178\times10^{-1}}$	$34.2028 + (-1.341896 \times 10^{1})C^{-4.448669\times10^{-2}}$
	DenseNet	$24.6690 + (-3.074005)N^{-5.708895\times10^{-1}}$	$2041.8949 + (-2.019874 \times 10^{3})C^{-2.314068\times10^{-4}}$
	EfficientNet	$23.7658 + (-9.452580 \times 10^{-1})N^{-6.994187\times10^{-1}}$	$1546.5953 + (-1.525287 \times 10^{3})C^{-3.317050\times10^{-4}}$
Desc. Text.	CNN	$22.2887 + (1.268340 \times 10^{12})N^{-1.639200\times10^{1}}$	$3744.4447 + (-3.724119 \times 10^{3})C^{-1.082274\times10^{-4}}$
	ResNet	$63.1084 + (-4.146958 \times 10^{1})N^{-2.520233\times10^{-2}}$	$25.4812 + (-5.954346)C^{-4.455387\times10^{-1}}$
	VGG	$39.3413 + (-1.678618 \times 10^{1})N^{-3.735827\times10^{-2}}$	$25.1106 + (-1.287993 \times 10^{1})C^{-6.692755\times10^{-1}}$
	DenseNet	$28.4511 + (-9.335082)N^{-3.332191\times10^{-1}}$	$25.3320 + (-4.332469 \times 10^{1})C^{-1.359420}$
	EfficientNet	$25.8027 + (-4.733167)N^{-5.008198\times10^{-1}}$	$24.9894 + (-1.783990 \times 10^{1})C^{-7.818619\times10^{-1}}$
Flower	CNN	$46.0983 + (-2.108755 \times 10^{1})N^{-9.949239\times10^{-5}}$	$25.0367 + (-1.238336 \times 10^{2})C^{-2.624658}$
	ResNet	$2748.2716 + (-2.719645 \times 10^{3})N^{-1.315553\times10^{-4}}$	$7803.8907 + (-7.779494 \times 10^{3})C^{-1.517311\times10^{-4}}$
	VGG	$50.3590 + (-2.410098 \times 10^{1})N^{-2.888038\times10^{-2}}$	$12282.4605 + (-1.226111 \times 10^{4})C^{-1.269159\times10^{-4}}$
	DenseNet	$31.5939 + (-6.395017)N^{-6.065856\times10^{-1}}$	$49.9150 + (-2.426361 \times 10^{1})C^{-4.316837\times10^{-2}}$
	EfficientNet	$29.5391 + (-1.054914 \times 10^{1})N^{-1.432826}$	$5437.4314 + (-5.419027 \times 10^{3})C^{-4.443195\times10^{-4}}$
Fungi	CNN	$-4136.5243 + (4.162303 \times 10^{3})N^{-1.641488\times10^{-4}}$	$-1063.5257 + (1.087794 \times 10^{3})C^{-1.727301\times10^{-4}}$
	ResNet	$25.3766 + (-1.140253 \times 10^{1})N^{-1.671725}$	$25.8988 + (-7.093830)C^{-4.752622\times10^{-1}}$
	VGG	$25.2755 + (-2.787233 \times 10^{4})N^{-5.779705}$	$6303.8576 + (-6.283341 \times 10^{3})C^{-1.439196\times10^{-4}}$
	DenseNet	$2251.7400 + (-2.227038 \times 10^{3})N^{-1.300032\times10^{-4}}$	$8443.0963 + (-8.421292 \times 10^{3})C^{-1.012216\times10^{-4}}$
	EfficientNet	$4304.0620 + (-4.280076 \times 10^{3})N^{-8.789412\times10^{-5}}$	$10301.7185 + (-1.028190 \times 10^{4})C^{-1.117026\times10^{-4}}$
Omniglot	CNN	$-50.1671 + (7.230481 \times 10^{1})N^{-5.786245\times10^{-4}}$	$1491.4353 + (-1.470521 \times 10^{3})C^{-1.549913\times10^{-4}}$
	ResNet	$30.0316 + (-1.035817 \times 10^{2})N^{-2.859173}$	$4349.7122 + (-4.322788 \times 10^{3})C^{-1.511430\times10^{-4}}$
	VGG	$28.5829 + (-5.115559 \times 10^{13})N^{-1.728633\times10^{1}}$	$29.6921 + (-6.916130)C^{-3.796910\times10^{-1}}$
	DenseNet	$30.4855 + (-1.315538 \times 10^{15})N^{-1.913954\times10^{1}}$	$33.2497 + (-8.214019)C^{-2.412283\times10^{-1}}$
	EfficientNet	$30.9538 + (-5.270456)N^{-5.294971\times10^{-1}}$	$30.8949 + (-2.202916 \times 10^{1})C^{-6.232129\times10^{-1}}$
Traffic Sign	CNN		$3265.5145 + (-3.243101 \times 10^{3})C^{-1.109198\times10^{-4}}$
	ResNet		$29.0434 + (-2.601508 \times 10^{1})C^{-9.703780\times10^{-1}}$
	VGG		$40.1975 + (-2.139162 \times 10^{1})C^{-1.161814\times10^{-1}}$
	DenseNet		$97.1794 + (-7.395121 \times 10^{1})C^{-2.124478\times10^{-2}}$
	EfficientNet		$28.0005 + (-1.790353 \times 10^{2})C^{-1.525033}$
Average	CNN	$22.6665 + (3.089323)N^{-1.102054}$	$23.0262 + (-2.577530)C^{-6.548699\times10^{-1}}$
	ResNet	$26.7862 + (-3.063717)N^{-4.892585\times10^{-1}}$	$28.6708 + (-6.792174)C^{-2.449362\times10^{-1}}$
	VGG	$26.1893 + (-2.935863)N^{-4.386443\times10^{-1}}$	$7377.2497 + (-7.355976 \times 10^{3})C^{-1.315395\times10^{-4}}$
	DenseNet	$27.3191 + (-4.395526)N^{-5.600893\times10^{-1}}$	$1850.0645 + (-1.826800 \times 10^{3})C^{-4.502698\times10^{-4}}$
	EfficientNet	$26.7086 + (-3.013176)N^{-4.466167\times10^{-1}}$	$29.6746 + (-1.237645 \times 10^{1})C^{-2.743273\times10^{-1}}$

Table 6. 5-way 1-shot power laws for models trained on Describable Texture.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$21.4008 + (1.477696 \times 10^{1})N^{-1.736192}$	$2521.3834 + (-2.501021 \times 10^{3})C^{-1.150771\times10^{-4}}$
	ResNet	$23.4427 + (-2.652488)N^{-1.600231\times10^{-1}}$	$21.9160 + (-1.076886 \times 10^{17})C^{-1.644649\times10^{1}}$
	VGG	$1384.7280 + (-1.363680 \times 10^{3})N^{-7.834595\times10^{-5}}$	$21.3355 + (-3.703994 \times 10^{18})C^{-1.747227\times10^{1}}$
	DenseNet	$22.0562 + (-4.203153 \times 10^{1})N^{-2.392815}$	$21.8662 + (-6.072810 \times 10^{-1})C^{-7.300234\times10^{-1}}$
	EfficientNet	$22.0887 + (-2.977386 \times 10^{1})N^{-2.086810}$	$22.1175 + (-9.980519)C^{-7.812849\times10^{-1}}$
Bird	CNN	$-1956.3783 + (1.979285 \times 10^{3})N^{-1.679876\times10^{-4}}$	$1012.2705 + (-9.910866 \times 10^{2})C^{-1.116567\times10^{-4}}$
	ResNet	$738.5653 + (-7.165168 \times 10^{2})N^{-3.109320\times10^{-4}}$	$22.6817 + (-1.135443 \times 10^{2})C^{-2.259176}$
	VGG	$522.6280 + (-5.014783 \times 10^{2})N^{-5.342870\times10^{-4}}$	$1311.9558 + (-1.291099 \times 10^{3})C^{-1.673223\times10^{-4}}$
	DenseNet	$23.0718 + (-2.518153)N^{-7.569310\times10^{-1}}$	$582.0913 + (-5.602041 \times 10^{2})C^{-3.766209\times10^{-4}}$
	EfficientNet	$22.6550 + (-5.516921 \times 10^{14})N^{-1.899318\times10^{1}}$	$5541.9943 + (-5.522418 \times 10^{3})C^{-1.183671\times10^{-4}}$
Flower	CNN	$-4776.0934 + (4.804721 \times 10^{3})N^{-1.783335\times10^{-4}}$	$2766.7276 + (-2.744164 \times 10^{3})C^{-2.704636\times10^{-4}}$
	ResNet	$29.6485 + (-3.828951)N^{-5.128747\times10^{-1}}$	$28.9021 + (-5.701136)C^{-8.040549\times10^{-1}}$
	VGG	$32.0942 + (-7.850959)N^{-1.167331\times10^{-1}}$	$26.3815 + (-3.585693)C^{-4.897144\times10^{-1}}$
	DenseNet	$28.9479 + (-1.211330 \times 10^{1})N^{-1.676966}$	$28.5681 + (-1.972755 \times 10^{18})C^{-1.679579\times10^{1}}$
	EfficientNet	$2080.0243 + (-2.053671 \times 10^{3})N^{-8.988686\times10^{-5}}$	$8416.1403 + (-8.395803 \times 10^{3})C^{-1.932498\times10^{-4}}$
Fungi	CNN	$23.6138 + (3.512594 \times 10^{1})N^{-2.220307}$	$23.7711 + (-7.017471 \times 10^{2})C^{-2.510297}$
	ResNet	$2095.3551 + (-2.071840 \times 10^{3})N^{-1.282875\times10^{-4}}$	$546.2164 + (-5.217474 \times 10^{2})C^{-1.327076\times10^{-4}}$
	VGG	$24.4688 + (-4.472832 \times 10^{14})N^{-1.835397\times10^{1}}$	$5691.1425 + (-5.670194 \times 10^{3})C^{-1.372638\times10^{-4}}$
	DenseNet	$2751.8761 + (-2.728559 \times 10^{3})N^{-1.236570\times10^{-4}}$	$24.3709 + (1.292324 \times 10^{18})C^{-1.716009\times10^{1}}$
	EfficientNet	$1714.6080 + (-1.691294 \times 10^{3})N^{-8.247866\times10^{-5}}$	$2487.6461 + (-2.466050 \times 10^{3})C^{-2.133550\times10^{-4}}$
Omniglot	CNN	$22.3849 + (-1.604549)N^{-1.363022}$	$22.2030 + (4.788491 \times 10^{1})C^{-2.177041}$
	ResNet	$28.5895 + (-8.628576)N^{-7.471480\times10^{-1}}$	$278.3521 + (-2.502240 \times 10^{2})C^{8.530419\times10^{-4}}$
	VGG	$27.3896 + (-1.101812 \times 10^{1})N^{-1.043821}$	$26.6105 + (-8.071776 \times 10^{19})C^{-1.810299\times10^{1}}$
	DenseNet	$30.4618 + (-6.824900)N^{-3.169660\times10^{-1}}$	$28.4368 + (-1.586305 \times 10^{19})C^{-1.786170\times10^{1}}$
	EfficientNet	$28.3741 + (-1.264829 \times 10^{2})N^{-2.759702}$	$36.1522 + (-1.534722 \times 10^{1})C^{-1.297309\times10^{-1}}$
Traffic Sign	CNN	$23.4912 + (2.293387)N^{-7.209053\times10^{-1}}$	$826.6624 + (-8.035835 \times 10^{2})C^{-1.358518\times10^{-4}}$
	ResNet	$3259.3266 + (-3.236409 \times 10^{3})N^{-1.916586\times10^{-4}}$	$24.5630 + (-8.519508 \times 10^{19})C^{-1.850999\times10^{1}}$
	VGG	$24.1001 + (-3.443083 \times 10^{1})N^{-1.916926}$	$3724.1619 + (-3.701800 \times 10^{3})C^{-1.107036\times10^{-4}}$
	DenseNet	$3053.1483 + (-3.029588 \times 10^{3})N^{-1.048868\times10^{-4}}$	$-119.7805 + (1.439783 \times 10^{2})C^{-1.645452\times10^{-4}}$
	EfficientNet	$30.4861 + (-7.106048)N^{-5.217235\times10^{-2}}$	$4997.8112 + (-4.977111 \times 10^{3})C^{-1.496973\times10^{-4}}$
Average	CNN	$-658.7934 + (6.829184 \times 10^{2})N^{-4.235662\times10^{-4}}$	$1805.8570 + (-1.784011 \times 10^{3})C^{-1.633285\times10^{-4}}$
	ResNet	$29.6684 + (-6.121536)N^{-8.859917\times10^{-2}}$	$24.9753 + (-4.887750 \times 10^{18})C^{-1.761671\times10^{1}}$
	VGG	$24.5720 + (-6.141343)N^{-9.597804\times10^{-1}}$	$1613.4802 + (-1.591057 \times 10^{3})C^{-2.335028\times10^{-4}}$
	DenseNet	$28.3053 + (-4.494976)N^{-1.028670\times10^{-1}}$	$25.0148 + (-4.028500 \times 10^{2})C^{-2.966862}$
	EfficientNet	$25.3098 + (-1.591382)N^{-2.818221\times10^{-1}}$	$4499.1455 + (-4.478534 \times 10^{3})C^{-1.984521\times10^{-4}}$

Table 7. 5-way 1-shot power laws for models trained on Flower.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$-1013.5895 + (1.035818 \times 10^{3})N^{-1.045772\times10^{-4}}$	$21.9381 + (-1.981000 \times 10^{3})C^{-3.286542}$
	ResNet	$21.7424 + (-2.722180)N^{-6.886557\times10^{-1}}$	$20.3360 + (1.738777)C^{-1.403230\times10^{-1}}$
	VGG	$21.4901 + (-7.561778)N^{-1.236343}$	$277.6417 + (-2.567460 \times 10^{2})C^{-2.421674\times10^{-4}}$
	DenseNet	$21.5006 + (-2.031501 \times 10^{2})N^{-2.567480}$	$-555.5253 + (5.773774 \times 10^{2})C^{-2.081392\times10^{-4}}$
	EfficientNet	$21.7191 + (-1.812361 \times 10^{5})N^{-5.224064}$	$21.4508 + (-5.384397 \times 10^{13})C^{-1.284869\times10^{1}}$
Bird	CNN	$22.0925 + (-1.544441 \times 10^{18})N^{-1.740066\times10^{1}}$	$22.2079 + (-1.488153 \times 10^{1})C^{-1.118477}$
	ResNet	$22.8921 + (-3.571599)N^{-5.792331\times10^{-1}}$	$-105.8048 + (1.280101 \times 10^{2})C^{-2.477879\times10^{-4}}$
	VGG	$1247.2912 + (-1.226326 \times 10^{3})N^{-2.562908\times10^{-4}}$	$22.0632 + (-4.131706)C^{-7.973438\times10^{-1}}$
	DenseNet	$2092.7892 + (-2.071515 \times 10^{3})N^{-1.242315\times10^{-4}}$	$18.0649 + (4.096209)C^{-3.983855\times10^{-3}}$
	EfficientNet	$2230.2572 + (-2.208041 \times 10^{3})N^{-1.553490\times10^{-4}}$	$23.3459 + (-1.455177 \times 10^{1})C^{-1.066801}$
Desc. Text.	CNN	$22.1826 + (-2.479874 \times 10^{18})N^{-1.763240\times10^{1}}$	$22.5689 + (-3.631614)C^{-4.686720\times10^{-1}}$
	ResNet	$1270.5387 + (-1.249114 \times 10^{3})N^{-1.663442\times10^{-4}}$	$21.9539 + (-3.424814 \times 10^{17})C^{-1.701602\times10^{1}}$
	VGG	$26.8439 + (-6.320726)N^{-7.080675\times10^{-2}}$	$21.6742 + (-2.705809 \times 10^{19})C^{-1.810098\times10^{1}}$
	DenseNet	$28.4370 + (-6.983501)N^{-2.796954\times10^{-2}}$	$22.0307 + (-1.201539 \times 10^{19})C^{-1.803230\times10^{1}}$
	EfficientNet	$2291.5396 + (-2.269971 \times 10^{3})N^{-1.215986\times10^{-4}}$	$22.3062 + (-8.557170 \times 10^{19})C^{-1.834374\times10^{1}}$
Fungi	CNN	$23.3680 + (-2.394768 \times 10^{6})N^{-6.191955}$	$23.4368 + (-2.255061 \times 10^{4})C^{-4.159534}$
	ResNet	$24.8112 + (-4.264096)N^{-7.522388\times10^{-1}}$	$4614.6490 + (-4.587643 \times 10^{3})C^{1.512505\times10^{-4}}$
	VGG	$24.0886 + (-1.604541 \times 10^{19})N^{-1.803749\times10^{1}}$	$23.7475 + (-1.196650 \times 10^{20})C^{-1.826884\times10^{1}}$
	DenseNet	$24.5669 + (-3.111204 \times 10^{18})N^{-1.746006\times10^{1}}$	$24.4300 + (-2.742916 \times 10^{19})C^{-1.779542\times10^{1}}$
	EfficientNet	$3578.0134 + (-3.554919 \times 10^{3})N^{-1.191744\times10^{-4}}$	$24.8647 + (-3.816923 \times 10^{1})C^{-1.390932}$
Omniglot	CNN	$-414.3203 + (4.379238 \times 10^{2})N^{-2.012069\times10^{-4}}$	$24.0895 + (-7.305899)C^{-5.212196\times10^{-1}}$
	ResNet	$30.9436 + (-9.217532)N^{-1.297477\times10^{-1}}$	$-4335.2610 + (4.363145 \times 10^{3})C^{-1.800532\times10^{-4}}$
	VGG	$4801.1540 + (-4.777318 \times 10^{3})N^{-1.245020\times10^{-4}}$	$25.6684 + (-7.111244 \times 10^{19})C^{-1.818383\times10^{1}}$
	DenseNet	$25.6667 + (-3.283946 \times 10^{18})N^{-1.714806\times10^{1}}$	$-3913.3039 + (3.941087 \times 10^{3})C^{-1.474499\times10^{-4}}$
	EfficientNet	$2368.4977 + (-2.344389 \times 10^{3})N^{-2.012745\times10^{-4}}$	$25.8474 + (-2.640285 \times 10^{18})C^{-1.716749\times10^{1}}$
Traffic Sign	CNN	$-4124.2732 + (4.151393 \times 10^{3})N^{-1.580323\times10^{-4}}$	$4573.5495 + (-4.551316 \times 10^{3})C^{-1.258844\times10^{-4}}$
	ResNet	$25.0305 + (-1.242847 \times 10^{19})N^{-1.781285\times10^{1}}$	$24.6342 + (-4.231688 \times 10^{19})C^{-1.814307\times10^{1}}$
	VGG	$24.4257 + (-9.025956 \times 10^{18})N^{-1.763091\times10^{1}}$	$252.8456 + (-2.290188 \times 10^{2})C^{-2.669047\times10^{-4}}$
	DenseNet	$-249.9544 + (2.747964 \times 10^{2})N^{-2.217738\times10^{-4}}$	$1546.7441 + (-1.522768 \times 10^{3})C^{-1.173805\times10^{-4}}$
	EfficientNet	$25.0385 + (-1.527837 \times 10^{19})N^{-1.784521\times10^{1}}$	$2405.8868 + (-2.383153 \times 10^{3})C^{-1.032180\times10^{-4}}$
Average	CNN	$-586.0343 + (6.093496 \times 10^{2})N^{-1.913071\times10^{-4}}$	$23.3776 + (-4.868016)C^{-5.562545\times10^{-1}}$
	ResNet	$24.0250 + (-3.400313)N^{-5.079056\times10^{-1}}$	$6.4133 + (1.772712 \times 10^{1})C^{-1.435753\times10^{-2}}$
	VGG	$23.9595 + (-2.227372)N^{-3.415867\times10^{-1}}$	$23.0228 + (-5.482410 \times 10^{20})C^{-1.922029\times10^{1}}$
	DenseNet	$23.4827 + (-7.027217 \times 10^{1})N^{-2.209200}$	$23.3847 + (-2.541803 \times 10^{18})C^{-1.748529\times10^{1}}$
	EfficientNet	$2656.6902 + (-2.633935 \times 10^{3})N^{-1.140875\times10^{-4}}$	$23.5806 + (-5.336627 \times 10^{1})C^{-1.760197}$

Table 8. 5-way 1-shot power laws for models trained on Fungi.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$-61.8120 + (8.350357 \times 10^1)N^{-7.699338\times10^{-4}}$	$21.6352 + (-1.102961 \times 10^{20})C^{-1.863243\times10^1}$
	ResNet	$22.7508 + (-1.563739)N^{-2.344455\times10^{-1}}$	$21.8150 + (-7.976878)C^{-1.924611}$
	VGG	$3148.1080 + (-3.127172 \times 10^3)N^{-8.973859\times10^{-5}}$	$21.6388 + (-1.728783 \times 10^{17})C^{-1.680439\times10^1}$
	DenseNet	$940.4213 + (-9.192035 \times 10^2)N^{-2.802262\times10^{-4}}$	$-926.2566 + (9.486473 \times 10^2)C^{-1.310169\times10^{-4}}$
	EfficientNet	$2234.8509 + (-2.213431 \times 10^3)N^{-8.342954\times10^{-5}}$	$21.9328 + (-2.105929 \times 10^{17})C^{-1.677314\times10^1}$
Bird	CNN	$-1580.2925 + (1.602929 \times 10^3)N^{-1.218963\times10^{-4}}$	$1157.6134 + (-1.136761 \times 10^3)C^{-2.575473\times10^{-4}}$
	ResNet	$8886.9851 + (-8.865314 \times 10^3)N^{-7.823261\times10^{-5}}$	$23.0313 + (6.647948)C^{-9.792309\times10^{-1}}$
	VGG	$10190.0017 + (-1.016951 \times 10^4)N^{-9.372112\times10^{-5}}$	$-139.8009 + (1.630744 \times 10^2)C^{-1.954904\times10^{-4}}$
	DenseNet	$8868.1100 + (-8.846979 \times 10^3)N^{-9.701722\times10^{-5}}$	$-1603.3930 + (1.627607 \times 10^3)C^{-1.458945\times10^{-4}}$
	EfficientNet	$4402.4127 + (-4.379537 \times 10^3)N^{-1.659970\times10^{-4}}$	$27.1547 + (-3.487227)C^{-6.542062\times10^{-2}}$
Desc. Text.	CNN	$472.2471 + (-4.504512 \times 10^2)N^{-1.434378\times10^{-4}}$	$572.5372 + (-5.508974 \times 10^2)C^{-1.086460\times10^{-4}}$
	ResNet	$4679.8630 + (-4.658469 \times 10^3)N^{-9.448095\times10^{-5}}$	$-169.8721 + (1.923645 \times 10^2)C^{-2.094978\times10^{-4}}$
	VGG	$5053.6787 + (-5.032649 \times 10^3)N^{-8.657719\times10^{-5}}$	$21.5693 + (2.609622)C^{-3.725027\times10^{-1}}$
	DenseNet	$5375.5515 + (-5.354219 \times 10^3)N^{-9.673946\times10^{-5}}$	$22.0340 + (1.276704)C^{-2.065462\times10^{-1}}$
	EfficientNet	$4168.5305 + (-4.146884 \times 10^3)N^{-1.022886\times10^{-4}}$	$22.6838 + (-3.650092 \times 10^{17})C^{-1.658990\times10^1}$
Flower	CNN	$-2128.3158 + (2.156354 \times 10^3)N^{-1.764488\times10^{-4}}$	$26.9297 + (-3.556391 \times 10^{20})C^{-1.863520\times10^1}$
	ResNet	$11736.4182 + (-1.170919 \times 10^4)N^{-8.412003\times10^{-5}}$	$-20.4878 + (5.138618 \times 10^1)C^{-7.223014\times10^{-3}}$
	VGG	$11574.5348 + (-1.154927 \times 10^4)N^{-1.074799\times10^{-4}}$	$-4126.7388 + (4.158407 \times 10^3)C^{-1.752010\times10^{-4}}$
	DenseNet	$4207.2471 + (-4.180424 \times 10^3)N^{-2.642393\times10^{-4}}$	$27.7950 + (4.226993)C^{-1.836036\times10^{-1}}$
	EfficientNet	$48.1695 + (-1.928197 \times 10^1)N^{-4.627078\times10^{-2}}$	$31.1041 + (-1.578097 \times 10^1)C^{-1.160511}$
Omniglot	CNN	$-1455.1954 + (1.479083 \times 10^3)N^{-1.314696\times10^{-4}}$	$1499.1575 + (-1.477715 \times 10^3)C^{-2.535445\times10^{-4}}$
	ResNet	$27.8038 + (-1.204446)N^{-6.847731\times10^{-1}}$	$27.3011 + (1.410233 \times 10^5)C^{-4.953103}$
	VGG	$26.9579 + (-1.602741 \times 10^{14})N^{-1.793500\times10^1}$	$26.4053 + (1.637853 \times 10^{17})C^{-1.606145\times10^1}$
	DenseNet	$27.9368 + (-3.793057 \times 10^{14})N^{-1.869632\times10^1}$	$-1037.1381 + (1.064832 \times 10^3)C^{-1.383823\times10^{-4}}$
	EfficientNet	$27.7259 + (-7.277565)N^{-1.009850}$	$1722.6470 + (-1.693463 \times 10^3)C^{3.381030\times10^{-4}}$
Traffic Sign	CNN	$-664.8319 + (6.886083 \times 10^2)N^{-1.328123\times10^{-4}}$	$23.7213 + (-7.525762 \times 10^{18})C^{-1.748740\times10^1}$
	ResNet	$-637.4883 + (6.630190 \times 10^2)N^{-1.474488\times10^{-4}}$	$-525.3205 + (5.511008 \times 10^2)C^{-4.254540\times10^{-4}}$
	VGG	$24.1615 + (5.824816)N^{-1.412371}$	$-371.9637 + (3.965616 \times 10^2)C^{-1.516544\times10^{-4}}$
	DenseNet	$1140.5595 + (-1.115969 \times 10^3)N^{-1.214338\times10^{-4}}$	$25.4734 + (-2.703782 \times 10^{17})C^{-1.591782\times10^1}$
	EfficientNet	$25.4062 + (-2.228803 \times 10^2)N^{-3.110611}$	$1061.2233 + (-1.036777 \times 10^3)C^{-1.538466\times10^{-4}}$
Average	CNN	$-1220.3992 + (1.244037 \times 10^3)N^{-1.153854\times10^{-4}}$	$23.1698 + (-8.575117 \times 10^{20})C^{-1.935716\times10^1}$
	ResNet	$2399.6151 + (-2.375512 \times 10^3)N^{-1.600591\times10^{-4}}$	$24.6571 + (2.824515)C^{-7.289026\times10^{-1}}$
	VGG	$4942.7853 + (-4.919666 \times 10^3)N^{-9.966580\times10^{-5}}$	$-1096.2748 + (1.121495 \times 10^3)C^{-1.788740\times10^{-4}}$
	DenseNet	$3670.9646 + (-3.647207 \times 10^3)N^{-1.367661\times10^{-4}}$	$-579.8446 + (6.053407 \times 10^2)C^{-1.997690\times10^{-4}}$
	EfficientNet	$26.8200 + (-3.768956)N^{-4.059386\times10^{-1}}$	$25.3304 + (-6.591492 \times 10^{17})C^{-1.712772\times10^1}$

Table 9. 5-way 1-shot power laws for models trained on ImageNet.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$21.9527 + (4.032354)N^{-1.720650}$	$21.9937 + (-5.908183 \times 10^{1})C^{-2.664139}$
	ResNet	$24.7525 + (-4.887383)N^{-8.057684\times10^{-1}}$	$23.4866 + (-6.313964)C^{-1.218383}$
	VGG	$5072.6125 + (-5.050229 \times 10^{3})N^{-1.064749\times10^{-4}}$	$577.7098 + (-5.552575 \times 10^{2})C^{-3.030502\times10^{-4}}$
	DenseNet	$25.8547 + (-4.024597)N^{-4.328134\times10^{-1}}$	$23.8428 + (-2.238344)C^{-7.975441\times10^{-1}}$
	EfficientNet	$4585.5179 + (-4.562189 \times 10^{3})N^{-1.042384\times10^{-4}}$	$761.4554 + (-7.393618 \times 10^{2})C^{-5.879906\times10^{-4}}$
Bird	CNN	$22.4351 + (4.528687 \times 10^{13})N^{-1.819681\times10^{1}}$	$27.3163 + (-6.258136)C^{-5.372594\times10^{-2}}$
	ResNet	$31.2686 + (-1.593026 \times 10^{1})N^{-7.082612\times10^{-1}}$	$27.4540 + (-4.534274)C^{-5.227035\times10^{-1}}$
	VGG	$21329.1088 + (-2.130586 \times 10^{4})N^{-7.717369\times10^{-5}}$	$26.2777 + (-8.434298 \times 10^{1})C^{-2.389636}$
	DenseNet	$32.2037 + (-1.823397 \times 10^{1})N^{-6.724366\times10^{-1}}$	$27.2787 + (-2.303607 \times 10^{1})C^{-1.586695}$
	EfficientNet	$35.2471 + (-1.149896 \times 10^{1})N^{-3.105254\times10^{-1}}$	$29.5642 + (-1.067349 \times 10^{1})C^{-6.023568\times10^{-1}}$
Desc. Text.	CNN	$23.0019 + (-6.270164 \times 10^{-1})N^{-4.938939\times10^{-1}}$	$22.9599 + (-3.527562)C^{-6.368657\times10^{-1}}$
	ResNet	$31.5451 + (-1.085290 \times 10^{1})N^{-2.438237\times10^{-1}}$	$-231.5486 + (2.569453 \times 10^{2})C^{-7.528249\times10^{-4}}$
	VGG	$27.8328 + (-8.389718)N^{-4.175246\times10^{-1}}$	$-1448.0396 + (1.473074 \times 10^{3})C^{-1.579117\times10^{-4}}$
	DenseNet	$8910.7143 + (-8.887891 \times 10^{3})N^{-1.343205\times10^{-4}}$	$23.0327 + (2.307317)C^{-4.263146\times10^{-2}}$
	EfficientNet	$29.7882 + (-7.290413)N^{-3.012974\times10^{-1}}$	$26.4840 + (-3.008027)C^{-3.020153\times10^{-1}}$
Flower	CNN	$26.2461 + (1.063648 \times 10^{1})N^{-1.690550}$	$26.2091 + (-2.431189 \times 10^{1})C^{-1.100457}$
	ResNet	$33.5939 + (-2.346974 \times 10^{1})N^{-1.500594}$	$32.6416 + (-8.809410)C^{-6.252300\times10^{-1}}$
	VGG	$34.3834 + (-6.138511)N^{-3.612986\times10^{-1}}$	$32.4095 + (-1.003879 \times 10^{1})C^{-4.897461\times10^{-1}}$
	DenseNet	$34.8740 + (-2.181235 \times 10^{1})N^{-1.269916}$	$33.2510 + (-6.650087)C^{-9.344070\times10^{-1}}$
	EfficientNet	$41.8339 + (-9.124385)N^{-7.405766\times10^{-2}}$	$34.5950 + (-1.699957 \times 10^{1})C^{-6.773284\times10^{-1}}$
Fungi	CNN	$-509.0488 + (5.336427 \times 10^{2})N^{-1.317507\times10^{-4}}$	$4201.7572 + (-4.178639 \times 10^{3})C^{-1.041800\times10^{-4}}$
	ResNet	$4959.9449 + (-4.933754 \times 10^{3})N^{-1.080449\times10^{-4}}$	$27.1443 + (-1.850574 \times 10^{1})C^{-1.271786}$
	VGG	$1128.0062 + (-1.103165 \times 10^{3})N^{-6.440241\times10^{-4}}$	$-1304.8173 + (1.331769 \times 10^{3})C^{-1.428852\times10^{-4}}$
	DenseNet	$29.8733 + (-8.258523)N^{-6.195409\times10^{-1}}$	$1598.6287 + (-1.572558 \times 10^{3})C^{-2.516926\times10^{-4}}$
	EfficientNet	$30.5395 + (-9.522299)N^{-6.046982\times10^{-1}}$	$27.8205 + (-1.954263 \times 10^{1})C^{-1.036721}$
Omniglot	CNN	$13.6233 + (9.590289)N^{-2.496333\times10^{-2}}$	$1185.6785 + (-1.164422 \times 10^{3})C^{-2.728125\times10^{-4}}$
	ResNet	$-917.2796 + (9.485936 \times 10^{2})N^{-1.201057\times10^{-4}}$	$30.5368 + (-2.986414 \times 10^{15})C^{-1.909384\times10^{1}}$
	VGG	$532.4766 + (-5.025621 \times 10^{2})N^{-9.973952\times10^{-5}}$	$30.0470 + (-1.053342 \times 10^{1})C^{-8.186476\times10^{-1}}$
	DenseNet	$31.5737 + (-5.863492)N^{-1.058741}$	$31.3224 + (-7.655773)C^{-1.177433}$
	EfficientNet	$32.5133 + (-4.596161)N^{-3.903997\times10^{-1}}$	$30.8011 + (-5.676328 \times 10^{1})C^{-1.948547}$
Traffic Sign	CNN	$23.3643 + (8.183132 \times 10^{-1})N^{-8.609143\times10^{-1}}$	$414.8651 + (-3.915976 \times 10^{2})C^{-1.759953\times10^{-4}}$
	ResNet	$27.8207 + (-1.098491 \times 10^{15})N^{-1.877626\times10^{1}}$	$977.9131 + (-9.511503 \times 10^{2})C^{-6.859245\times10^{-5}}$
	VGG	$38.5606 + (-1.514105 \times 10^{1})N^{-6.576825\times10^{-2}}$	$25.3418 + (-2.431566 \times 10^{13})C^{-1.757895\times10^{1}}$
	DenseNet	$28.7497 + (-2.234716 \times 10^{1})N^{-1.499904}$	$27.6150 + (1.669298 \times 10^{14})C^{-1.838063\times10^{1}}$
	EfficientNet	$28.5729 + (-8.151357 \times 10^{2})N^{-3.621350}$	$28.6562 + (-4.267122)C^{-3.456202\times10^{-1}}$
Average	CNN	$23.3005 + (7.239823 \times 10^{-1})N^{-5.923027\times10^{-1}}$	$28.6845 + (-6.925951)C^{-6.719183\times10^{-2}}$
	ResNet	$29.1717 + (-1.005010 \times 10^{1})N^{-9.681091\times10^{-1}}$	$27.4479 + (-8.841685)C^{-1.145929}$
	VGG	$2410.7552 + (-2.385545 \times 10^{3})N^{-3.251683\times10^{-4}}$	$26.5193 + (-2.136315 \times 10^{1})C^{-1.785428}$
	DenseNet	$30.3308 + (-8.680429)N^{-6.543338\times10^{-1}}$	$28.1386 + (-2.376950)C^{-6.356347\times10^{-1}}$
	EfficientNet	$31.4265 + (-6.114191)N^{-3.570699\times10^{-1}}$	$28.7223 + (-9.183906)C^{-7.190499\times10^{-1}}$

Table 10. 5-way 1-shot power laws for models trained on Omniglot.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$22.1855 + (-8.255858 \times 10^{12})N^{-1.691991\times10^{1}}$	$-504.1393 + (5.266011 \times 10^{2})C^{-1.348534\times10^{-4}}$
	ResNet	$21.5689 + (-1.670906 \times 10^{2})N^{-3.784077}$	$984.1171 + (-9.631427 \times 10^{2})C^{-1.210300\times10^{-4}}$
	VGG	$200.4047 + (-1.789152 \times 10^{2})N^{-2.899000\times10^{-4}}$	$21.8347 + (-2.289338)C^{-7.582400\times10^{-1}}$
	DenseNet	$195.1308 + (-1.739715 \times 10^{2})N^{-1.737360\times10^{-4}}$	$21.1386 + (3.530019 \times 10^{17})C^{-1.693686\times10^{1}}$
	EfficientNet	$21.6745 + (-6.281038 \times 10^{13})N^{-1.783153\times10^{1}}$	$1385.9419 + (-1.365073 \times 10^{3})C^{-1.299693\times10^{-4}}$
Bird	CNN	$302.2999 + (-2.813750 \times 10^{2})N^{-3.737334\times10^{-4}}$	$18.8315 + (3.259656)C^{-7.701476\times10^{-2}}$
	ResNet	$21.5147 + (-1.360150 \times 10^{1})N^{-1.850518}$	$21.3780 + (7.418711 \times 10^{18})C^{-1.797038\times10^{1}}$
	VGG	$21.5412 + (-5.542021 \times 10^{-1})N^{-4.288866\times10^{-1}}$	$-327.5751 + (3.491608 \times 10^{2})C^{-1.425434\times10^{-4}}$
	DenseNet	$618.0090 + (-5.970225 \times 10^{2})N^{-2.004957\times10^{-4}}$	$21.2530 + (8.163471 \times 10^{-1})C^{-7.586665\times10^{-1}}$
	EfficientNet	$21.3251 + (-3.051647 \times 10^{2})N^{-4.026550}$	$1334.5699 + (-1.313980 \times 10^{3})C^{-1.075284\times10^{-4}}$
Desc. Text.	CNN	$288.4171 + (-2.672170 \times 10^{2})N^{-3.249479\times10^{-4}}$	$21.2436 + (2.383391)C^{-6.323313\times10^{-1}}$
	ResNet	$21.4658 + (-5.719724 \times 10^{1})N^{-2.653235}$	$122.5239 + (-1.013058 \times 10^{2})C^{-2.792074\times10^{-4}}$
	VGG	$21.5739 + (-1.610857)N^{-8.016755\times10^{-1}}$	$21.4111 + (-1.598342 \times 10^{1})C^{-1.949075}$
	DenseNet	$21.2917 + (-2.298817)N^{-2.306144}$	$150.6318 + (-1.294629 \times 10^{2})C^{-1.264447\times10^{-4}}$
	EfficientNet	$-496.0266 + (5.175999 \times 10^{2})N^{-1.140273\times10^{-4}}$	$2223.2252 + (-2.202819 \times 10^{3})C^{-9.722955\times10^{-5}}$
Flower	CNN	$1100.8731 + (-1.079017 \times 10^{3})N^{-2.993880\times10^{-4}}$	$-721.4543 + (7.464779 \times 10^{2})C^{-7.480797\times10^{-4}}$
	ResNet	$23.2675 + (-1.835358 \times 10^{1})N^{-1.550920}$	$677.1082 + (-6.549660 \times 10^{2})C^{-2.642267\times10^{-4}}$
	VGG	$23.1599 + (-5.216170 \times 10^{-1})N^{-1.292140}$	$23.2390 + (-1.795963 \times 10^{1})C^{-1.625997}$
	DenseNet	$22.9619 + (9.189006 \times 10^{12})N^{-1.735229\times10^{1}}$	$183.3423 + (-1.611673 \times 10^{2})C^{-8.004284\times10^{-4}}$
	EfficientNet	$23.1869 + (-1.118787 \times 10^{15})N^{-1.894438\times10^{1}}$	$2097.9988 + (-2.075802 \times 10^{3})C^{-1.145368\times10^{-4}}$
Fungi	CNN	$23.5616 + (-1.803638)N^{-7.720326\times10^{-1}}$	$1927.1763 + (-1.905151 \times 10^{3})C^{-1.719647\times10^{-4}}$
	ResNet	$1818.1846 + (-1.796463 \times 10^{3})N^{-1.620186\times10^{-4}}$	$22.8102 + (-1.052173 \times 10^{1})C^{-1.484746}$
	VGG	$23.4877 + (-9.098563 \times 10^{-1})N^{-1.994974\times10^{-1}}$	$22.9285 + (8.109522)C^{-1.096841}$
	DenseNet	$-1028.7258 + (1.052869 \times 10^{3})N^{-2.227612\times10^{-4}}$	$4304.9926 + (-4.283969 \times 10^{3})C^{-1.240806\times10^{-4}}$
	EfficientNet	$1201.7479 + (-1.179162 \times 10^{3})N^{-1.058872\times10^{-4}}$	$22.8699 + (-1.468439)C^{-7.625592\times10^{-1}}$
Traffic Sign	CNN	$24.1888 + (-6.048420)N^{-1.314642}$	$24.0750 + (-6.566354 \times 10^{-1})C^{-5.309326\times10^{-1}}$
	ResNet	$22.6008 + (3.380327)N^{-5.018885\times10^{-1}}$	$488.2746 + (-4.650343 \times 10^{2})C^{-1.592791\times10^{-4}}$
	VGG	$18.7286 + (4.535255)N^{-2.172602\times10^{-2}}$	$23.3017 + (-1.178468 \times 10^{18})C^{-1.696248\times10^{1}}$
	DenseNet	$310.5421 + (-2.877609 \times 10^{2})N^{-1.033639\times10^{-4}}$	$23.0396 + (-2.954740 \times 10^{1})C^{-1.673535}$
	EfficientNet	$2786.4381 + (-2.761772 \times 10^{3})N^{1.421775\times10^{-4}}$	$2803.8826 + (-2.781803 \times 10^{3})C^{-1.193661\times10^{-4}}$
Average	CNN	$22.8654 + (-1.382106)N^{-4.654900\times10^{-1}}$	$22.2861 + (1.027071)C^{-4.052982\times10^{-1}}$
	ResNet	$22.2880 + (-1.250950 \times 10^{1})N^{-2.024917}$	$354.8976 + (-3.329668 \times 10^{2})C^{-2.130141\times10^{-4}}$
	VGG	$22.3126 + (-6.547201 \times 10^{-1})N^{-8.919293\times10^{-1}}$	$22.3446 + (-1.973230 \times 10^{2})C^{-3.144683}$
	DenseNet	$22.1770 + (2.163673 \times 10^{12})N^{-1.682737\times10^{1}}$	$1150.4262 + (-1.128883 \times 10^{3})C^{-1.195235\times10^{-4}}$
	EfficientNet	$22.2896 + (-2.292461 \times 10^{12})N^{-1.696799\times10^{1}}$	$1818.1922 + (-1.796754 \times 10^{3})C^{-1.092170\times10^{-4}}$

Table 11. 5-way 1-shot power laws for models trained on Quickdraw.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$-517.9344 + (5.405973 \times 10^{2})N^{-1.084520\times10^{-4}}$	$22.4932 + (-2.327263 \times 10^{1})C^{-1.948100}$
	ResNet	$22.2905 + (-7.379860 \times 10^{13})N^{-1.826834\times10^{1}}$	$22.3436 + (-6.169113 \times 10^{2})C^{-3.008023}$
	EfficientNet	$23.4018 + (-1.037999)N^{-1.377843\times10^{-2}}$	$22.2239 + (-1.329396 \times 10^{2})C^{-2.172371}$
Bird	CNN	$21.3754 + (-5.234673 \times 10^{11})N^{-1.626573\times10^{1}}$	$177.4266 + (-1.562193 \times 10^{2})C^{-1.387392\times10^{-4}}$
	ResNet	$-111.8211 + (1.335280 \times 10^{2})N^{-8.716999\times10^{-5}}$	$-197.9263 + (2.198047 \times 10^{2})C^{-1.661106\times10^{-4}}$
	EfficientNet	$21.2279 + (2.419962)N^{-1.374197}$	$21.2820 + (-1.085860 \times 10^{19})C^{-1.809031\times10^{1}}$
Desc. Text.	CNN	$-95.8753 + (1.177243 \times 10^{2})N^{-1.785211\times10^{-4}}$	$899.0336 + (-8.777396 \times 10^{2})C^{-1.269591\times10^{-4}}$
	ResNet	$22.1700 + (-1.298827)N^{-1.037377}$	$22.1152 + (-9.499713 \times 10^{17})C^{-1.733066\times10^{1}}$
	EfficientNet	$552.1555 + (-5.305047 \times 10^{2})N^{-8.366413\times10^{-5}}$	$21.7689 + (-6.390258)C^{-1.273033}$
Flower	CNN	$-737.1876 + (7.609078 \times 10^{2})N^{-1.246065\times10^{-4}}$	$559.8849 + (-5.370574 \times 10^{2})C^{-2.451801\times10^{-4}}$
	ResNet	$23.9830 + (2.021044 \times 10^{-1})N^{-2.200328\times10^{-1}}$	$571.7003 + (-5.489002 \times 10^{2})C^{-4.767497\times10^{-4}}$
	EfficientNet	$232.5096 + (-2.089153 \times 10^{2})N^{-1.114662\times10^{-4}}$	$914.0955 + (-8.917345 \times 10^{2})C^{-3.887125\times10^{-4}}$
Fungi	CNN	$1124.5556 + (-1.101750 \times 10^{3})N^{-1.236083\times10^{-4}}$	$288.7987 + (-2.660203 \times 10^{2})C^{-8.885589\times10^{-5}}$
	ResNet	$23.7749 + (-3.867799 \times 10^{1})N^{-2.000267}$	$23.6341 + (-8.989051 \times 10^{17})C^{-1.691786\times10^{1}}$
	EfficientNet	$23.8747 + (-8.997046)N^{-1.281655}$	$23.3280 + (4.346526 \times 10^{-1})C^{-3.992645\times10^{-1}}$
Omniglot	CNN	$29.6968 + (-8.865645 \times 10^{-2})N^{-9.461015\times10^{-1}}$	$29.2860 + (-6.590613 \times 10^{18})C^{-1.757900\times10^{1}}$
	ResNet	$35.4697 + (-5.234097 \times 10^{13})N^{-1.747353\times10^{1}}$	$38.0387 + (-1.488931 \times 10^{1})C^{-4.295154\times10^{-1}}$
	EfficientNet	$2041.2644 + (-2.011049 \times 10^{3})N^{-2.682897\times10^{-4}}$	$31.4092 + (-4.215291 \times 10^{1})C^{-1.432046\times10^{1}}$
Traffic Sign	CNN	$-1331.2010 + (1.354858 \times 10^{3})N^{-1.479156\times10^{-4}}$	$23.0952 + (1.440079 \times 10^{1})C^{-1.664601}$
	ResNet	$23.0150 + (-1.734250 \times 10^{6})N^{-9.041734}$	$23.3272 + (-1.669551 \times 10^{19})C^{-1.777146\times10^{1}}$
	EfficientNet	$1100.4549 + (-1.077441 \times 10^{3})N^{-1.075360\times10^{-4}}$	$627.8951 + (-6.054485 \times 10^{2})C^{-3.372467\times10^{-4}}$
Average	CNN	$-287.8010 + (3.114913 \times 10^{2})N^{-1.212626\times10^{-4}}$	$158.6631 + (-1.353724 \times 10^{2})C^{-2.693402\times10^{-4}}$
	ResNet	$24.6226 + (-5.052239 \times 10^{13})N^{-1.788042\times10^{1}}$	$24.6029 + (-3.029611 \times 10^{5})C^{-5.233488}$
	EfficientNet	$623.0652 + (-5.994944 \times 10^{2})N^{-2.289442\times10^{-4}}$	$24.2201 + (-1.733882 \times 10^{1})C^{-1.142022}$

Table 12. 5-way 1-shot power laws for models trained on Traffic Sign.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$22.4313 + (-2.875613 \times 10^{11})N^{-1.540155\times10^{1}}$	$22.1277 + (1.880510 \times 10^{1})C^{-1.854599}$
	ResNet	$21.7581 + (4.521949)N^{-1.764239}$	$21.6697 + (-2.398420 \times 10^{2})C^{-2.376125}$
	VGG	$21.8999 + (-6.440277)N^{-1.548939}$	$21.4458 + (-7.831996 \times 10^{19})C^{-1.861199\times10^{1}}$
	DenseNet	$1156.5580 + (-1.135332 \times 10^{3})N^{-1.261595\times10^{-4}}$	$735.3227 + (-7.143030 \times 10^{2})C^{-1.428143\times10^{-4}}$
	EfficientNet	$21.5117 + (-9.504050)N^{-1.940557}$	$21.2551 + (-1.322052 \times 10^{3})C^{-3.132739}$
Bird	CNN	$20.7740 + (2.656098)N^{-1.267202\times10^{-1}}$	$22.7369 + (-4.275229)C^{-1.163715}$
	ResNet	$885.3548 + (-8.625366 \times 10^{2})N^{4.224574\times10^{-4}}$	$4069.2050 + (-4.049924 \times 10^{3})C^{-1.582720\times10^{-4}}$
	VGG	$21.4971 + (3.400523)N^{-3.312063\times10^{-1}}$	$31.9829 + (-1.192992 \times 10^{1})C^{-6.662281\times10^{-2}}$
	DenseNet	$22.4889 + (8.213971 \times 10^{-1})N^{-4.553152\times10^{-1}}$	$547.7674 + (-5.263233 \times 10^{2})C^{-4.668607\times10^{-4}}$
	EfficientNet	$20.8405 + (5.495706)N^{-9.818418\times10^{-1}}$	$2760.7555 + (-2.741085 \times 10^{3})C^{-1.490839\times10^{-4}}$
Desc. Text.	CNN	$21.6858 + (6.230802 \times 10^{-1})N^{-5.287328\times10^{-1}}$	$21.9398 + (-4.589069 \times 10^{18})C^{-1.766907\times10^{1}}$
	ResNet	$21.0804 + (1.056599)N^{-3.726021\times10^{-1}}$	$21.8959 + (-2.051110)C^{-3.285377\times10^{-1}}$
	VGG	$-242.1144 + (2.639497 \times 10^{2})N^{-2.436452\times10^{-4}}$	$1143.4110 + (-1.122624 \times 10^{3})C^{-1.805155\times10^{-4}}$
	DenseNet	$21.7055 + (4.129308 \times 10^{1})N^{-3.385343}$	$405.5817 + (-3.842233 \times 10^{2})C^{-1.549279\times10^{-4}}$
	EfficientNet	$21.0205 + (3.738157)N^{-1.942626}$	$21.5398 + (-4.379763)C^{-5.487443\times10^{-1}}$
Flower	CNN	$-1042.4456 + (1.069888 \times 10^{3})N^{-2.532309\times10^{-4}}$	$27.2333 + (-7.287856 \times 10^{1})C^{-1.837032}$
	ResNet	$4262.5748 + (-4.234508 \times 10^{3})N^{2.142271\times10^{-4}}$	$6259.0283 + (-6.236223 \times 10^{3})C^{-1.065067\times10^{-4}}$
	VGG	$-4091.5953 + (4.119747 \times 10^{3})N^{-1.193123\times10^{-4}}$	$798.3369 + (-7.741089 \times 10^{2})C^{-7.444491\times10^{-4}}$
	DenseNet	$2151.4569 + (-2.122811 \times 10^{3})N^{2.362427\times10^{-4}}$	$42.4123 + (-1.830384 \times 10^{1})C^{-4.036260\times10^{-2}}$
	EfficientNet	$2145.1498 + (-2.120637 \times 10^{3})N^{1.395785\times10^{-4}}$	$27.6073 + (-7.950666)C^{-1.805966\times10^{-1}}$
Fungi	CNN	$24.2030 + (-2.070501)N^{-8.133759\times10^{-1}}$	$28.6118 + (-5.576920)C^{-7.172830\times10^{-2}}$
	ResNet	$23.3461 + (-3.108899 \times 10^{-1})N^{3.537573\times10^{-1}}$	$23.1353 + (-2.908511 \times 10^{18})C^{-1.703436\times10^{1}}$
	VGG	$-1037.8053 + (1.061593 \times 10^{3})N^{-1.424367\times10^{-4}}$	$321.1718 + (-2.972955 \times 10^{2})C^{-1.309126\times10^{-4}}$
	DenseNet	$1982.3567 + (-1.958304 \times 10^{3})N^{1.323084\times10^{-4}}$	$23.9155 + (-3.796143)C^{-5.044975\times10^{-1}}$
	EfficientNet	$22.2640 + (7.914834 \times 10^{-1})N^{-6.988680\times10^{-1}}$	$2560.1214 + (-2.538676 \times 10^{3})C^{-1.311935\times10^{-4}}$
Omniglot	CNN	$2507.1322 + (-2.474984 \times 10^{3})N^{-1.715902\times10^{-4}}$	$32.7439 + (1.639647 \times 10^{17})C^{-1.657223\times10^{1}}$
	ResNet	$4330.3022 + (-4.303382 \times 10^{3})N^{-7.783207\times10^{-5}}$	$29.3597 + (-1.652127 \times 10^{20})C^{-1.813379\times10^{1}}$
	VGG	$4271.6750 + (-4.238421 \times 10^{3})N^{1.489645\times10^{-4}}$	$33.1038 + (-3.636528 \times 10^{1})C^{-8.316413\times10^{-1}}$
	DenseNet	$1147.4873 + (-1.114353 \times 10^{3})N^{4.008023\times10^{-4}}$	$32.3912 + (-1.208480 \times 10^{17})C^{-1.553281\times10^{1}}$
	EfficientNet	$32.6335 + (-7.503372)N^{-1.807338\times10^{-1}}$	$4460.6028 + (-4.440804 \times 10^{3})C^{-3.982684\times10^{-4}}$
Average	CNN	$25.1589 + (-1.222365 \times 10^{-1})N^{-6.874215\times10^{-1}}$	$25.2676 + (-2.797865)C^{-9.058078\times10^{-1}}$
	ResNet	$2189.1869 + (-2.165012 \times 10^{3})N^{1.127400\times10^{-4}}$	$23.8844 + (-6.510815 \times 10^{1})C^{-1.678416}$
	VGG	$20.4875 + (5.007987)N^{-6.646033\times10^{-2}}$	$26.7490 + (-5.357408)C^{-2.391384\times10^{-1}}$
	DenseNet	$24.3182 + (2.071385)N^{-5.971643\times10^{-1}}$	$39.2008 + (-1.588846 \times 10^{1})C^{-2.240662\times10^{-2}}$
	EfficientNet	$353.4983 + (-3.305953 \times 10^{2})N^{-8.887010\times10^{-5}}$	$2070.6324 + (-2.050386 \times 10^{3})C^{-3.137885\times10^{-4}}$

Table 13. 5-way 5-shot power laws for models trained on Aircraft.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Bird	CNN	$1297.4912 + (-1.273371 \times 10^3)N^{-1.274682\times10^{-4}}$	$666.2228 + (-6.419596 \times 10^2)C^{-9.762661\times10^{-5}}$
	ResNet	$26.0436 + (-3.279345)N^{-1.309061\times10^{-1}}$	$23.7799 + (-8.360675 \times 10^2)C^{-2.701859}$
	VGG	$24.3871 + (-1.399672 \times 10^{14})N^{-1.788778\times10^{1}}$	$24.6864 + (-5.843194 \times 10^1)C^{-1.228925}$
	DenseNet	$2704.2978 + (-2.681306 \times 10^3)N^{-1.160894\times10^{-4}}$	$24.0559 + (-4.029547 \times 10^4)C^{-4.112868}$
	EfficientNet	$-1356.9419 + (1.381860 \times 10^3)N^{-1.632340\times10^{-4}}$	$24.5006 + (-5.708626 \times 10^1)C^{-1.337902}$
Desc. Text.	CNN	$4947.7036 + (-4.925446 \times 10^3)N^{-9.708745\times10^{-5}}$	$22.8761 + (1.423786)C^{-2.409649\times10^{-1}}$
	ResNet	$7530.5471 + (-7.508310 \times 10^3)N^{-6.537464\times10^{-5}}$	$28.4630 + (-6.904602)C^{-7.395635\times10^{-2}}$
	VGG	$4508.7729 + (-4.486635 \times 10^3)N^{-1.016531\times10^{-4}}$	$23.2847 + (-6.578498 \times 10^2)C^{-2.413282}$
	DenseNet	$4280.8510 + (-4.258285 \times 10^3)N^{-7.994246\times10^{-5}}$	$23.6802 + (-8.267235 \times 10^1)C^{-1.919042}$
	EfficientNet	$3745.3509 + (-3.723638 \times 10^3)N^{-1.790773\times10^{-4}}$	$23.3731 + (-6.836805 \times 10^{16})C^{-1.619361\times10^{1}}$
Flower	CNN	$3321.8902 + (-3.291653 \times 10^3)N^{-8.113857\times10^{-5}}$	$287.7630 + (-2.570696 \times 10^2)C^{1.708925\times10^{-4}}$
	ResNet	$5855.3508 + (-5.827676 \times 10^3)N^{-1.358289\times10^{-4}}$	$29.5623 + (-1.207254 \times 10^3)C^{-2.411821}$
	VGG	$3296.4799 + (-3.269214 \times 10^3)N^{-1.978214\times10^{-4}}$	$6203.2748 + (-6.178240 \times 10^3)C^{-1.297232\times10^{-4}}$
	DenseNet	$230.3178 + (-2.000215 \times 10^2)N^{1.292688\times10^{-4}}$	$30.4858 + (-1.642357 \times 10^3)C^{-2.508243}$
	EfficientNet	$1907.7917 + (-1.878678 \times 10^3)N^{-8.144369\times10^{-5}}$	$40.4180 + (-1.647169 \times 10^1)C^{-9.388727\times10^{-2}}$
Fungi	CNN	$3746.6203 + (-3.721847 \times 10^3)N^{-1.177836\times10^{-4}}$	$26.1057 + (-3.808154 \times 10^{18})C^{-1.754449\times10^{1}}$
	ResNet	$25.6927 + (-2.460379)N^{-7.700576\times10^{-1}}$	$25.7735 + (-3.713390 \times 10^2)C^{-2.045052}$
	VGG	$25.3704 + (-4.250546 \times 10^{13})N^{-1.737639\times10^{1}}$	$24.5974 + (-1.654562 \times 10^{20})C^{-1.850250\times10^{1}}$
	DenseNet	$259.7964 + (-2.343776 \times 10^2)N^{-9.128774\times10^{-5}}$	$25.8979 + (-1.613564 \times 10^{17})C^{-1.556244\times10^{1}}$
	EfficientNet	$25.1611 + (-1.177149 \times 10^6)N^{-8.052523}$	$4264.2451 + (-4.241667 \times 10^3)C^{-1.102128\times10^{-4}}$
Omniglot	CNN	$3085.9647 + (-3.054019 \times 10^3)N^{-1.096199\times10^{-4}}$	$622.0896 + (-5.901175 \times 10^2)C^{-1.336904\times10^{-4}}$
	ResNet	$38.2967 + (-2.712363 \times 10^{14})N^{-1.766422\times10^{1}}$	$37.5778 + (-5.627546 \times 10^{18})C^{-1.718324\times10^{1}}$
	VGG	$6446.8239 + (-6.412689 \times 10^3)N^{-1.053207\times10^{-4}}$	$34.5783 + (-4.260470 \times 10^1)C^{-1.684413\times10^{1}}$
	DenseNet	$43.5899 + (-1.456047 \times 10^1)N^{-2.358821\times10^{-1}}$	$36.9955 + (6.931143 \times 10^{11})C^{-1.118009\times10^{1}}$
	EfficientNet	$908.8504 + (-8.716935 \times 10^2)N^{-1.313335\times10^{-4}}$	$37.2245 + (-1.255145 \times 10^{17})C^{-1.519317\times10^{1}}$
Traffic Sign	CNN	$-1223.1703 + (1.252350 \times 10^3)N^{-2.130144\times10^{-4}}$	$29.0238 + (-1.718042 \times 10^1)C^{-7.966717\times10^{-1}}$
	ResNet	$26.4483 + (-7.213567 \times 10^{-1})N^{-9.373895\times10^{-1}}$	$26.6234 + (-1.773821 \times 10^{20})C^{-1.836847\times10^{1}}$
	VGG	$26.6987 + (-2.294150 \times 10^5)N^{-6.462187}$	$32.1638 + (-1.381406 \times 10^1)C^{-2.101356\times10^{-1}}$
	DenseNet	$26.7461 + (-1.070670 \times 10^1)N^{-1.654988}$	$6333.2922 + (-6.311916 \times 10^3)C^{-1.818875\times10^{-4}}$
	EfficientNet	$25.9361 + (-4.040152 \times 10^{14})N^{-1.834103\times10^{1}}$	$26.4692 + (-5.056216 \times 10^1)C^{-1.607208}$
Average	CNN	$2365.5872 + (-2.338502 \times 10^3)N^{-1.007843\times10^{-4}}$	$1199.4507 + (-1.172503 \times 10^3)C^{-1.177500\times10^{-4}}$
	ResNet	$32.2563 + (-5.927035)N^{-9.896593\times10^{-2}}$	$27.7283 + (-1.002488 \times 10^4)C^{-3.527760}$
	VGG	$1435.4414 + (-1.409378 \times 10^3)N^{-3.203963\times10^{-4}}$	$27.6735 + (-2.513558 \times 10^1)C^{-9.196737\times10^{-1}}$
	DenseNet	$4174.8785 + (-4.148419 \times 10^3)N^{-9.721504\times10^{-5}}$	$28.1345 + (-2.861051 \times 10^1)C^{-1.166762}$
	EfficientNet	$27.8309 + (-2.458915)N^{-9.618302\times10^{-1}}$	$27.7133 + (-3.476294 \times 10^1)C^{-1.252658}$

Table 14. 5-way 5-shot power laws for models trained on Bird.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$-141.9000 + (1.665801 \times 10^{2})N^{-2.292470\times10^{-4}}$	$25.6153 + (-4.177415)C^{-3.529170\times10^{-1}}$
	ResNet	$7199.8861 + (-7.178146 \times 10^{3})N^{-8.424268\times10^{-5}}$	$-280.0107 + (3.039255 \times 10^{2})C^{-4.628571\times10^{-4}}$
	VGG	$31.7144 + (-1.022218 \times 10^{1})N^{-6.412606\times10^{-2}}$	$23.0955 + (-1.793756 \times 10^{19})C^{-1.811765\times10^{1}}$
	DenseNet	$8684.7905 + (-8.663431 \times 10^{3})N^{-8.598243\times10^{-5}}$	$-257.3588 + (2.817949 \times 10^{2})C^{-9.554964\times10^{-4}}$
	EfficientNet	$4404.7457 + (-4.382589 \times 10^{3})N^{-9.618141\times10^{-5}}$	$287.7212 + (-2.649100 \times 10^{2})C^{-1.716192\times10^{-4}}$
Desc. Text.	CNN	$25.5940 + (-2.857723)N^{-3.362603\times10^{-1}}$	$24.2072 + (9.675255 \times 10^{4})C^{-5.084016}$
	ResNet	$40.8243 + (-1.774654 \times 10^{1})N^{-3.555274\times10^{-2}}$	$24.6889 + (-1.780578 \times 10^{5})C^{-5.479579}$
	VGG	$4764.3860 + (-4.742167 \times 10^{3})N^{-1.585847\times10^{-4}}$	$24.2564 + (-2.447894 \times 10^{16})C^{-1.537744\times10^{1}}$
	DenseNet	$9938.1641 + (-9.915439 \times 10^{3})N^{-7.760340\times10^{-5}}$	$-516.2087 + (5.413903 \times 10^{2})C^{-1.596834\times10^{-4}}$
	EfficientNet	$25.8645 + (-4.929434)N^{-6.281741\times10^{-1}}$	$24.6993 + (-7.591741 \times 10^{2})C^{-2.663054}$
Flower	CNN	$30.7648 + (2.001565)N^{-1.172455\times10^{-1}}$	$4515.7134 + (-4.486434 \times 10^{3})C^{-1.400271\times10^{-4}}$
	ResNet	$34.3094 + (-1.050655 \times 10^{1})N^{-7.729486\times10^{-1}}$	$33.0548 + (-4.670120 \times 10^{1})C^{-1.601461}$
	VGG	$2275.5113 + (-2.245693 \times 10^{3})N^{-2.999253\times10^{-4}}$	$31.0978 + (-1.151630 \times 10^{1})C^{-1.159392}$
	DenseNet	$3359.9838 + (-3.329234 \times 10^{3})N^{-2.444650\times10^{-4}}$	$2383.1922 + (-2.351864 \times 10^{3})C^{-1.410199\times10^{-4}}$
	EfficientNet	$35.5050 + (-7.741681)N^{-4.919847\times10^{-1}}$	$33.6709 + (-2.751020 \times 10^{1})C^{-9.333027\times10^{-1}}$
Fungi	CNN	$26.9875 + (-6.788359 \times 10^{1})N^{-2.574910}$	$3371.5186 + (-3.346785 \times 10^{3})C^{-1.302671\times10^{-4}}$
	ResNet	$29.4000 + (-1.010967 \times 10^{1})N^{-7.202535\times10^{-1}}$	$28.0490 + (-5.481262 \times 10^{19})C^{-1.809585\times10^{1}}$
	VGG	$553.0375 + (-5.273331 \times 10^{2})N^{-1.405338\times10^{-3}}$	$27.5132 + (-1.427239 \times 10^{4})C^{-3.653616}$
	DenseNet	$30.2888 + (-1.237369 \times 10^{1})N^{-7.119437\times10^{-1}}$	$28.8678 + (-3.407405)C^{-3.152967\times10^{-1}}$
	EfficientNet	$35.6966 + (-1.004003 \times 10^{1})N^{-8.803543\times10^{-2}}$	$2545.9366 + (-2.519956 \times 10^{3})C^{-1.386126\times10^{-4}}$
Omniglot	CNN	$27.2856 + (-1.920872 \times 10^{-1})N^{-3.449643\times10^{-1}}$	$27.1733 + (-3.532837 \times 10^{19})C^{-1.821120\times10^{1}}$
	ResNet	$18110.3608 + (-1.808633 \times 10^{4})N^{-1.029315\times10^{-4}}$	$30.7378 + (-7.279123 \times 10^{-3})C^{1.259518}$
	VGG	$17567.4136 + (-1.754372 \times 10^{4})N^{-9.947736\times10^{-5}}$	$5357.8480 + (-5.323987 \times 10^{3})C^{2.368146\times10^{-4}}$
	DenseNet	$8846.8139 + (-8.822512 \times 10^{3})N^{-1.980963\times10^{-4}}$	$30.5398 + (-5.195418 \times 10^{-10})C^{4.783137}$
	EfficientNet	$40.6848 + (-1.682922 \times 10^{1})N^{-1.858227\times10^{-1}}$	$32.7532 + (-6.734283 \times 10^{-1})C^{2.703756\times10^{-1}}$
Traffic Sign	CNN	$1857.4999 + (-1.829184 \times 10^{3})N^{1.172446\times10^{-4}}$	$27.9559 + (-1.282661 \times 10^{20})C^{-1.829483\times10^{1}}$
	ResNet	$27.8142 + (-3.277057 \times 10^{1})N^{-1.954196}$	$294.8452 + (-2.672866 \times 10^{2})C^{-1.390797\times10^{-4}}$
	VGG	$27.1827 + (-4.416947 \times 10^{11})N^{-1.503808\times10^{1}}$	$4241.1001 + (-4.215690 \times 10^{3})C^{-1.051838\times10^{-4}}$
	DenseNet	$27.7569 + (-1.208051 \times 10^{2})N^{-2.507114}$	$27.5052 + (-7.771965 \times 10^{2})C^{-3.198767}$
	EfficientNet	$27.2674 + (-4.754167 \times 10^{11})N^{-1.462436\times10^{1}}$	$27.3716 + (-6.357018 \times 10^{2})C^{-2.328304}$
Average	CNN	$27.1739 + (-4.954493 \times 10^{11})N^{-1.616411\times10^{1}}$	$2388.7743 + (-2.363097 \times 10^{3})C^{-1.416872\times10^{-4}}$
	ResNet	$2243.8119 + (-2.218544 \times 10^{3})N^{-3.728391\times10^{-4}}$	$-1011.1155 + (1.039266 \times 10^{3})C^{-1.263837\times10^{-4}}$
	VGG	$4152.9775 + (-4.127928 \times 10^{3})N^{-1.781394\times10^{-4}}$	$26.9517 + (-5.175937 \times 10^{2})C^{-3.344368}$
	DenseNet	$36.3850 + (-1.176578 \times 10^{1})N^{-1.096650\times10^{-1}}$	$-454.3930 + (4.822765 \times 10^{2})C^{-1.193408\times10^{-4}}$
	EfficientNet	$29.9566 + (-6.266167)N^{-4.064145\times10^{-1}}$	$27.6957 + (-4.643539 \times 10^{2})C^{-2.540669}$

Table 15. 5-way 5-shot power laws for models trained on COCO.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$-288.7254 + (3.147442 \times 10^2)N^{-1.200802\times10^{-4}}$	$4342.6158 + (-4.319236 \times 10^3)C^{-1.302065\times10^{-4}}$
	ResNet	$25.7624 + (-9.704097)N^{-1.327674}$	$25.7344 + (-3.661110 \times 10^1)C^{-1.138454}$
	VGG	$25.6750 + (-3.011819)N^{-4.787600\times10^{-1}}$	$25.3694 + (-2.001733 \times 10^1)C^{-8.005768\times10^{-1}}$
	DenseNet	$26.4250 + (-3.155604 \times 10^1)N^{-1.859316}$	$27.0834 + (-1.453339 \times 10^1)C^{-5.844777\times10^{-1}}$
	EfficientNet	$26.1698 + (-1.586017 \times 10^1)N^{-1.291501}$	$9362.3523 + (-9.344210 \times 10^3)C^{-1.751228\times10^{-4}}$
Bird	CNN	$25.7250 + (8.249795 \times 10^{13})N^{-1.842800\times10^{1}}$	$3914.0095 + (-3.890308 \times 10^3)C^{-1.067454\times10^{-4}}$
	ResNet	$27.4584 + (-4.100241)N^{-1.073380}$	$7586.8794 + (-7.563038 \times 10^3)C^{-1.102179\times10^{-4}}$
	VGG	$27.4119 + (-1.869527)N^{-2.342922\times10^{-1}}$	$33.4721 + (-1.311205 \times 10^1)C^{-1.433390\times10^{-1}}$
	DenseNet	$28.2018 + (-4.210334)N^{-8.522246\times10^{-1}}$	$7132.6425 + (-7.108524 \times 10^3)C^{-1.131081\times10^{-4}}$
	EfficientNet	$27.6498 + (-1.340329)N^{-3.041176\times10^{-1}}$	$10051.4405 + (-1.002980 \times 10^4)C^{-1.197403\times10^{-4}}$
Desc. Text.	CNN	$27.6789 + (-2.238824)N^{-3.143267\times10^{-1}}$	$8619.1813 + (-8.597294 \times 10^3)C^{-1.242387\times10^{-4}}$
	ResNet	$34.3027 + (-1.190748 \times 10^1)N^{-3.343765\times10^{-1}}$	$29.9246 + (-3.260115 \times 10^1)C^{-9.977888\times10^{-1}}$
	VGG	$33.4755 + (-9.135646)N^{-2.926858\times10^{-1}}$	$32.9211 + (-2.130016 \times 10^1)C^{-4.305449\times10^{-1}}$
	DenseNet	$33.3616 + (-1.160998 \times 10^1)N^{-4.538316\times10^{-1}}$	$30.7980 + (-3.351990 \times 10^1)C^{-9.782515\times10^{-1}}$
	EfficientNet	$31.7223 + (-8.529606)N^{-4.160320\times10^{-1}}$	$30.0050 + (-2.893338 \times 10^1)C^{-6.918675\times10^{-1}}$
Flower	CNN	$32.4109 + (-5.826179 \times 10^{-1})N^{-1.108480\times10^{-1}}$	$7272.1029 + (-7.244297 \times 10^3)C^{-1.216142\times10^{-4}}$
	ResNet	$38.8239 + (-5.193336)N^{-1.172992\times10^{-1}}$	$37.1781 + (-1.889504 \times 10^1)C^{-5.033954\times10^{-1}}$
	VGG	$35.0214 + (-7.198957)N^{-7.159702\times10^{-1}}$	$62.4761 + (-3.902488 \times 10^1)C^{-6.740991\times10^{-2}}$
	DenseNet	$36.8755 + (-4.867297)N^{-5.208351\times10^{-1}}$	$48.2131 + (-2.050834 \times 10^1)C^{-1.088867\times10^{-1}}$
	EfficientNet	$35.5161 + (-2.392761 \times 10^1)N^{-1.530384}$	$60.0864 + (-4.556970 \times 10^1)C^{-1.345697\times10^{-1}}$
Fungi	CNN	$222.9596 + (-1.964989 \times 10^2)N^{2.650055\times10^{-4}}$	$38.7173 + (-1.583685 \times 10^1)C^{-5.323138\times10^{-2}}$
	ResNet	$28.9326 + (-1.141224 \times 10^1)N^{-1.836743}$	$8605.1262 + (-8.581638 \times 10^3)C^{-1.309807\times10^{-4}}$
	VGG	$28.7770 + (-1.131711)N^{-5.335946\times10^{-1}}$	$14797.5886 + (-1.477752 \times 10^4)C^{-1.230347\times10^{-4}}$
	DenseNet	$29.8270 + (-3.491028 \times 10^1)N^{-1.805525}$	$30.5401 + (-1.505529 \times 10^1)C^{-4.987251\times10^{-1}}$
	EfficientNet	$28.8230 + (-1.268728 \times 10^1)N^{-1.258145}$	$18126.1018 + (-1.810678 \times 10^4)C^{-1.146586\times10^{-4}}$
Omniglot	CNN	$26.6733 + (1.602743 \times 10^{14})N^{-1.869102\times10^{1}}$	$4121.5347 + (-4.097471 \times 10^3)C^{-1.500690\times10^{-4}}$
	ResNet	$34.6894 + (-9.908674)N^{-1.516711}$	$34.1222 + (-5.953367 \times 10^{20})C^{-1.879437\times10^{1}}$
	VGG	$34.3207 + (-6.038002)N^{-5.896604\times10^{-1}}$	$39.2390 + (-1.501647 \times 10^1)C^{-2.132872\times10^{-1}}$
	DenseNet	$34.8783 + (-1.062753 \times 10^{14})N^{-1.814471\times10^{1}}$	$6138.3967 + (-6.107279 \times 10^3)C^{-1.402210\times10^{-4}}$
	EfficientNet	$34.9887 + (-5.825551 \times 10^1)N^{-1.918735}$	$36.0549 + (-4.376420 \times 10^1)C^{-7.290340\times10^{-1}}$
Traffic Sign	CNN		$3795.0800 + (-3.767112 \times 10^3)C^{-1.087130\times10^{-4}}$
	ResNet		$32.8440 + (-2.863138 \times 10^1)C^{-1.013732\times10^{1}}$
	VGG		$7786.5557 + (-7.764825 \times 10^3)C^{-3.083132\times10^{-4}}$
	DenseNet		$8997.7937 + (-8.972322 \times 10^3)C^{-2.406810\times10^{-4}}$
	EfficientNet		$37.4388 + (-3.259288 \times 10^1)C^{-4.676954\times10^{-1}}$
Average	CNN	$102.4018 + (-7.523832 \times 10^1)N^{-3.526374\times10^{-4}}$	$4660.7393 + (-4.636179 \times 10^3)C^{-1.432635\times10^{-4}}$
	ResNet	$30.9243 + (-4.865521)N^{-6.628912\times10^{-1}}$	$31.5201 + (-1.500461 \times 10^1)C^{-6.317713\times10^{-1}}$
	VGG	$30.5919 + (-4.397867)N^{-4.724602\times10^{-1}}$	$106.8143 + (-8.482329 \times 10^1)C^{-2.181935\times10^{-2}}$
	DenseNet	$31.3059 + (-9.189520)N^{-1.006532}$	$78.0787 + (-5.271297 \times 10^1)C^{-2.663537\times10^{-2}}$
	EfficientNet	$30.5877 + (-1.136807 \times 10^1)N^{-1.085713}$	$39.8681 + (-2.328099 \times 10^1)C^{-2.015300\times10^{-1}}$

Table 16. 5-way 5-shot power laws for models trained on Describable Texture.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$419.9637 + (-3.954281 \times 10^2)N^{-2.894110\times10^{-4}}$	$3048.1547 + (-3.025206 \times 10^3)C^{-1.344733\times10^{-4}}$
	ResNet	$25.3184 + (-4.476238)N^{-2.888187\times10^{-1}}$	$23.6204 + (-4.298798 \times 10^{17})C^{-1.665998\times10^{1}}$
	VGG	$3619.0052 + (-3.597493 \times 10^3)N^{-1.166151\times10^{-4}}$	$22.6848 + (-2.048747 \times 10^1)C^{-1.430886}$
	DenseNet	$24.6936 + (-4.058497)N^{-4.563243\times10^{-1}}$	$23.4957 + (-2.728219 \times 10^3)C^{-4.320480}$
	EfficientNet	$24.2975 + (-8.677756)N^{-9.758394\times10^{-1}}$	$23.9519 + (-2.690993 \times 10^1)C^{-9.403026\times10^{-1}}$
Bird	CNN	$-227.2083 + (2.526624 \times 10^2)N^{-7.254117\times10^{-4}}$	$2238.0866 + (-2.214729 \times 10^3)C^{-1.307158\times10^{-4}}$
	ResNet	$709.1606 + (-6.851613 \times 10^2)N^{-7.613813\times10^{-4}}$	$25.5277 + (-1.097674 \times 10^1)C^{-1.050696}$
	VGG	$25.9161 + (-7.144271)N^{-5.768795\times10^{-1}}$	$5401.9179 + (-5.381258 \times 10^3)C^{-1.494951\times10^{-4}}$
	DenseNet	$26.0859 + (-1.419494 \times 10^1)N^{-1.282900}$	$29.7078 + (-6.196770)C^{-9.393779\times10^{-2}}$
	EfficientNet	$25.3580 + (-1.105011 \times 10^{15})N^{-1.898988\times10^{1}}$	$9463.8720 + (-9.444795 \times 10^3)C^{-1.425727\times10^{-4}}$
Flower	CNN	$-1034.4753 + (1.066857 \times 10^3)N^{-1.705037\times10^{-4}}$	$6343.1540 + (-6.315309 \times 10^3)C^{-1.444887\times10^{-4}}$
	ResNet	$35.2592 + (-6.595663)N^{-3.849705\times10^{-1}}$	$33.0878 + (-5.378409 \times 10^2)C^{-2.439033}$
	VGG	$33.0879 + (-1.159980 \times 10^1)N^{-6.016866\times10^{-1}}$	$4922.1002 + (-4.895522 \times 10^3)C^{-1.636129\times10^{-4}}$
	DenseNet	$33.9602 + (-8.888621)N^{-8.378855\times10^{-1}}$	$32.8149 + (-4.535743 \times 10^{20})C^{-1.881778\times10^{1}}$
	EfficientNet	$6264.1096 + (-6.234390 \times 10^3)N^{-8.088830\times10^{-5}}$	$78.2007 + (-5.808280 \times 10^1)C^{-5.052216\times10^{-2}}$
Fungi	CNN	$26.7767 + (-5.897156)N^{-1.327137}$	$4014.3737 + (-3.990427 \times 10^3)C^{-1.204205\times10^{-4}}$
	ResNet	$31.5133 + (-6.880133)N^{-1.624105\times10^{-1}}$	$27.2001 + (-4.490627 \times 10^2)C^{-2.466810}$
	VGG	$27.7093 + (-1.814517 \times 10^1)N^{-1.080451}$	$26.8935 + (-8.688309)C^{-4.534608\times10^{-1}}$
	DenseNet	$27.5958 + (-4.415100)N^{-7.470619\times10^{-1}}$	$26.8286 + (-3.009063 \times 10^2)C^{-2.402644}$
	EfficientNet	$27.4992 + (-3.265967)N^{-2.212233\times10^{-1}}$	$5317.0462 + (-5.296020 \times 10^3)C^{-2.176771\times10^{-4}}$
Omniglot	CNN	$26.7835 + (-8.215753 \times 10^3)N^{-5.285537}$	$2604.5687 + (-2.579195 \times 10^3)C^{-1.204844\times10^{-4}}$
	ResNet	$32.3281 + (-1.555097 \times 10^1)N^{-8.935485\times10^{-1}}$	$30.7600 + (-5.952248 \times 10^{18})C^{-1.733169\times10^{1}}$
	VGG	$30.9373 + (-1.092058 \times 10^2)N^{-1.963767}$	$29.4111 + (-3.460841 \times 10^{19})C^{-1.772424\times10^{1}}$
	DenseNet	$32.9166 + (-3.985859 \times 10^1)N^{-1.286241}$	$2149.7206 + (-2.119998 \times 10^3)C^{-2.892802\times10^{-4}}$
	EfficientNet	$32.7882 + (-2.217253 \times 10^2)N^{-2.492145}$	$11362.6837 + (-1.133885 \times 10^4)C^{-1.462250\times10^{-4}}$
Traffic Sign	CNN	$-2119.3383 + (2.147201 \times 10^3)N^{-1.360051\times10^{-4}}$	$924.5749 + (-8.985925 \times 10^2)C^{-1.281024\times10^{-4}}$
	ResNet	$10380.4100 + (-1.035579 \times 10^4)N^{-8.829057\times10^{-5}}$	$26.7006 + (-9.836955 \times 10^{18})C^{-1.754653\times10^{1}}$
	VGG	$27.5461 + (-9.016931)N^{-8.546473\times10^{-1}}$	$3334.5170 + (-3.311904 \times 10^3)C^{-2.474428\times10^{-4}}$
	DenseNet	$8786.7825 + (-8.762223 \times 10^3)N^{-9.374674\times10^{-5}}$	$26.6498 + (-1.987947 \times 10^{19})C^{-1.796269\times10^{1}}$
	EfficientNet	$8972.5546 + (-8.947793 \times 10^3)N^{-7.948231\times10^{-5}}$	$3306.7501 + (-3.286624 \times 10^3)C^{-4.134681\times10^{-4}}$
Average	CNN	$-524.6277 + (5.517804 \times 10^2)N^{-1.073463\times10^{-4}}$	$4139.7325 + (-4.114823 \times 10^3)C^{-1.018643\times10^{-4}}$
	ResNet	$33.1411 + (-7.996767)N^{-1.443131\times10^{-1}}$	$27.7542 + (-7.018198 \times 10^{19})C^{-1.827763\times10^{1}}$
	VGG	$28.1231 + (-1.081861 \times 10^1)N^{-8.066729\times10^{-1}}$	$53.6831 + (-3.020718 \times 10^1)C^{-2.303869\times10^{-2}}$
	DenseNet	$29.2343 + (-6.817953)N^{-5.985583\times10^{-1}}$	$27.8823 + (-2.589779 \times 10^{17})C^{-1.606903\times10^{1}}$
	EfficientNet	$28.3760 + (-5.317283)N^{-6.439393\times10^{-1}}$	$8394.3419 + (-8.373659 \times 10^3)C^{-1.788056\times10^{-4}}$

Table 17. 5-way 5-shot power laws for models trained on Flower.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$24.4983 + (-6.034926 \times 10^3)N^{-3.750935}$	$24.1405 + (-2.639254 \times 10^1)C^{-2.214618}$
	ResNet	$23.4051 + (-7.199215)N^{-7.743851\times10^{-1}}$	$-829.1498 + (8.525202 \times 10^2)C^{-2.462750\times10^{-4}}$
	VGG	$23.1144 + (-1.430609 \times 10^1)N^{-1.136673}$	$1101.9944 + (-1.080486 \times 10^3)C^{-1.925837\times10^{-4}}$
	DenseNet	$23.2395 + (-6.141096)N^{-7.544130\times10^{-1}}$	$22.6372 + (-9.436387 \times 10^{17})C^{-1.704830\times10^1}$
	EfficientNet	$23.2856 + (-2.489297 \times 10^2)N^{-2.404592}$	$22.7751 + (-1.068963 \times 10^{20})C^{-1.839325\times10^1}$
Bird	CNN	$25.1040 + (-5.724297 \times 10^{19})N^{-1.825903\times10^1}$	$24.8317 + (-7.619272)C^{-7.836073\times10^{-1}}$
	ResNet	$26.0171 + (-1.553337 \times 10^1)N^{-8.085721\times10^{-1}}$	$-556.2540 + (5.808489 \times 10^2)C^{-1.982919\times10^{-4}}$
	VGG	$47.6400 + (-2.590620 \times 10^1)N^{-3.011064\times10^{-2}}$	$23.9078 + (-6.465713 \times 10^1)C^{-1.596292}$
	DenseNet	$2106.2412 + (-2.083740 \times 10^3)N^{-2.917827\times10^{-4}}$	$24.2608 + (-1.622828 \times 10^{19})C^{-1.810291\times10^1}$
	EfficientNet	$5189.8522 + (-5.165571 \times 10^3)N^{-1.310142\times10^{-4}}$	$26.5408 + (-1.631954 \times 10^1)C^{-7.805272\times10^{-1}}$
Desc. Text.	CNN	$26.6284 + (-4.261850)N^{-2.103567\times10^{-1}}$	$24.1958 + (-7.364087 \times 10^1)C^{-1.796824}$
	ResNet	$30.9790 + (-8.545837)N^{-8.179443\times10^{-2}}$	$24.0485 + (-1.413680 \times 10^{19})C^{-1.817200\times10^1}$
	VGG	$3340.3365 + (-3.318527 \times 10^3)N^{-1.995190\times10^{-4}}$	$23.5432 + (-1.005277 \times 10^2)C^{-1.822133}$
	DenseNet	$4494.2548 + (-4.471509 \times 10^3)N^{-1.075364\times10^{-4}}$	$24.1287 + (-7.257917 \times 10^{19})C^{-1.845376\times10^1}$
	EfficientNet	$4543.0340 + (-4.519854 \times 10^3)N^{-1.051496\times10^{-4}}$	$24.3484 + (-3.455353 \times 10^5)C^{-5.005936}$
Fungi	CNN	$2231.3357 + (-2.206444 \times 10^3)N^{-1.326269\times10^{-4}}$	$25.6625 + (-1.483342 \times 10^3)C^{-2.796045}$
	ResNet	$36.0317 + (-1.228896 \times 10^1)N^{-9.323506\times10^{-2}}$	$26.1751 + (-6.894193 \times 10^{17})C^{-1.688058\times10^1}$
	VGG	$2128.4155 + (-2.103673 \times 10^3)N^{-1.985633\times10^{-4}}$	$26.1524 + (-9.589407)C^{-7.201475\times10^{-1}}$
	DenseNet	$27.3046 + (-4.126006)N^{-5.275003\times10^{-1}}$	$26.0588 + (1.497565)C^{-8.492918\times10^{-1}}$
	EfficientNet	$30.4605 + (-7.676073)N^{-2.731416\times10^{-1}}$	$26.7992 + (-3.676224 \times 10^{20})C^{-1.857764\times10^1}$
Omniglot	CNN	$26.8356 + (3.995653 \times 10^{20})N^{-1.900326\times10^1}$	$27.3511 + (-1.932315 \times 10^1)C^{-1.106281}$
	ResNet	$29.3932 + (-1.242443 \times 10^1)N^{-4.781911\times10^{-1}}$	$28.3226 + (-2.410692 \times 10^{-4})C^{2.035236}$
	VGG	$2457.3862 + (-2.431391 \times 10^3)N^{-1.812844\times10^{-4}}$	$-1939.9795 + (1.968485 \times 10^3)C^{-1.300045\times10^{-4}}$
	DenseNet	$27.3973 + (-4.177853 \times 10^1)N^{-1.531405}$	$-6872.4644 + (6.904096 \times 10^3)C^{-1.510757\times10^{-4}}$
	EfficientNet	$8516.3035 + (-8.491332 \times 10^3)N^{-1.273672\times10^{-4}}$	$1515.3611 + (-1.488298 \times 10^3)C^{-2.884998\times10^{-4}}$
Traffic Sign	CNN	$-716.5020 + (7.444439 \times 10^2)N^{-2.157036\times10^{-4}}$	$27.3155 + (-6.934067 \times 10^3)C^{-3.319365}$
	ResNet	$27.5076 + (-1.256586 \times 10^{20})N^{-1.839225\times10^1}$	$26.5421 + (-8.152818 \times 10^{16})C^{-1.566081\times10^1}$
	VGG	$26.3002 + (7.317995 \times 10^1)N^{-1.993224}$	$26.4072 + (-3.932189 \times 10^{-1})C^{-1.060093\times10^{-1}}$
	DenseNet	$26.7907 + (-9.430989 \times 10^{18})N^{-1.782253\times10^1}$	$26.5486 + (2.361662 \times 10^{18})C^{-1.723067\times10^1}$
	EfficientNet	$27.0085 + (-4.079779 \times 10^{19})N^{-1.819722\times10^1}$	$26.5975 + (-1.191692 \times 10^1)C^{-9.356216\times10^{-1}}$
Average	CNN	$167.7855 + (-1.425506 \times 10^2)N^{-9.212852\times10^{-4}}$	$25.5444 + (-9.899723 \times 10^1)C^{-1.838971}$
	ResNet	$26.7056 + (-9.411316)N^{-6.471799\times10^{-1}}$	$-1062.8165 + (1.088637 \times 10^3)C^{-1.800620\times10^{-4}}$
	VGG	$2978.0437 + (-2.954203 \times 10^3)N^{-1.346209\times10^{-4}}$	$24.8434 + (-1.620507 \times 10^2)C^{-2.229689}$
	DenseNet	$26.4127 + (-3.817618)N^{-3.791090\times10^{-1}}$	$-789.5398 + (8.154432 \times 10^2)C^{-2.246335\times10^{-4}}$
	EfficientNet	$1656.0103 + (-1.631676 \times 10^3)N^{-3.409611\times10^{-4}}$	$25.8294 + (-1.321278 \times 10^3)C^{-2.808197}$

Table 18. 5-way 5-shot power laws for models trained on Fungi.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$1048.8391 + (-1.024599 \times 10^3)N^{-7.577124\times10^{-5}}$	$24.0651 + (1.157617 \times 10^1)C^{-1.297945}$
	ResNet	$25.6373 + (-3.364322)N^{-2.464625\times10^{-1}}$	$23.8286 + (-1.802580 \times 10^{18})C^{-1.735563\times10^1}$
	VGG	$7103.3703 + (-7.081203 \times 10^3)N^{-8.288747\times10^{-5}}$	$-279.7351 + (3.037225 \times 10^2)C^{-3.295853\times10^{-4}}$
	DenseNet	$47.5168 + (-2.510248 \times 10^1)N^{-2.322155\times10^{-2}}$	$-1084.7089 + (1.109712 \times 10^3)C^{-2.446415\times10^{-4}}$
	EfficientNet	$4279.5392 + (-4.256768 \times 10^3)N^{-1.080376\times10^{-4}}$	$-1084.7089 + (1.109712 \times 10^3)C^{-2.446415\times10^{-4}}$
Bird	CNN	$1146.2045 + (-1.121255 \times 10^3)N^{-1.393000\times10^{-4}}$	$25.2205 + (6.485634 \times 10^{19})C^{-1.842883\times10^1}$
	ResNet	$16174.0888 + (-1.615007 \times 10^4)N^{-7.822741\times10^{-5}}$	$26.9701 + (1.300880 \times 10^1)C^{-1.460678}$
	VGG	$14621.0163 + (-1.459802 \times 10^4)N^{-9.696723\times10^{-5}}$	$25.7035 + (2.203037)C^{-2.174798\times10^{-1}}$
	DenseNet	$14853.7296 + (-1.483005 \times 10^4)N^{-1.004811\times10^{-4}}$	$25.8994 + (3.346649)C^{-2.357951\times10^{-1}}$
	EfficientNet	$8452.2518 + (-8.426173 \times 10^3)N^{-1.543693\times10^{-4}}$	$25.8994 + (3.346649)C^{-2.357951\times10^{-1}}$
Desc. Text.	CNN	$4227.9427 + (-4.203771 \times 10^3)N^{-9.415461\times10^{-5}}$	$25.1070 + (8.400043 \times 10^3)C^{-4.166527}$
	ResNet	$9729.3563 + (-9.706024 \times 10^3)N^{-1.183601\times10^{-4}}$	$26.0840 + (-1.773679 \times 10^{18})C^{-1.737809\times10^1}$
	VGG	$11594.0527 + (-1.157102 \times 10^4)N^{-9.367341\times10^{-5}}$	$16.7799 + (1.126265 \times 10^1)C^{-5.433485\times10^{-2}}$
	DenseNet	$51.6847 + (-2.819446 \times 10^1)N^{-4.741904\times10^{-2}}$	$25.6773 + (2.240939)C^{-2.076061\times10^{-1}}$
	EfficientNet	$8999.6648 + (-8.975853 \times 10^3)N^{-1.087995\times10^{-4}}$	$25.6773 + (2.240939)C^{-2.076061\times10^{-1}}$
Flower	CNN	$32.2961 + (2.043626 \times 10^{12})N^{-1.605620\times10^1}$	$32.0699 + (-7.896660 \times 10^2)C^{-3.063597}$
	ResNet	$8231.4953 + (-8.199254 \times 10^3)N^{-1.089550\times10^{-4}}$	$-76.7792 + (1.114228 \times 10^2)C^{-5.377998\times10^{-4}}$
	VGG	$9636.1311 + (-9.605955 \times 10^3)N^{-1.268907\times10^{-4}}$	$-4766.6124 + (4.802756 \times 10^3)C^{-1.416636\times10^{-4}}$
	DenseNet	$8024.8652 + (-7.993249 \times 10^3)N^{-1.398092\times10^{-4}}$	$-425.5206 + (4.609477 \times 10^2)C^{-5.374504\times10^{-4}}$
	EfficientNet	$37.7279 + (-8.326541)N^{-5.570509\times10^{-1}}$	$-425.5206 + (4.609477 \times 10^2)C^{-5.374504\times10^{-4}}$
Omniglot	CNN	$27.0825 + (3.708209 \times 10^3)N^{-5.088314}$	$-183.2565 + (2.106285 \times 10^2)C^{-2.557761\times10^{-4}}$
	ResNet	$3399.8795 + (-3.370335 \times 10^3)N^{-9.162112\times10^{-5}}$	$30.1675 + (1.208584 \times 10^3)C^{-2.984275}$
	VGG	$31.4674 + (-5.354049)N^{-3.217118\times10^{-1}}$	$-2807.7779 + (2.838514 \times 10^3)C^{-1.606901\times10^{-4}}$
	DenseNet	$2238.6541 + (-2.209921 \times 10^3)N^{-2.835795\times10^{-4}}$	$-1015.6830 + (1.046122 \times 10^3)C^{-1.490128\times10^{-4}}$
	EfficientNet	$31.8990 + (-9.118800)N^{-6.077862\times10^{-1}}$	$-1015.6830 + (1.046122 \times 10^3)C^{-1.490128\times10^{-4}}$
Traffic Sign	CNN	$26.9189 + (2.020565)N^{-6.280902\times10^{-1}}$	$27.0149 + (-6.799339 \times 10^{19})C^{-1.835639\times10^1}$
	ResNet	$28.0603 + (6.965753 \times 10^{-1})N^{-1.153486}$	$28.0070 + (-6.727333 \times 10^{18})C^{-1.748502\times10^1}$
	VGG	$26.7226 + (3.260301 \times 10^{13})N^{-1.804978\times10^1}$	$3103.9410 + (-3.075337 \times 10^3)C^{1.474250\times10^{-4}}$
	DenseNet	$1738.9067 + (-1.710834 \times 10^3)N^{-8.840498\times10^{-5}}$	$28.0370 + (-6.665906 \times 10^1)C^{-2.217273}$
	EfficientNet	$28.2578 + (-3.943254 \times 10^2)N^{-3.932872}$	$28.0370 + (-6.665906 \times 10^1)C^{-2.217273}$
Average	CNN	$635.9138 + (-6.090684 \times 10^2)N^{-9.100933\times10^{-5}}$	$-84.5402 + (1.114398 \times 10^2)C^{-2.622178\times10^{-4}}$
	ResNet	$6526.6486 + (-6.499964 \times 10^3)N^{-1.021334\times10^{-4}}$	$28.0165 + (8.373106 \times 10^{-1})C^{-3.478768\times10^{-1}}$
	VGG	$10068.1409 + (-1.004269 \times 10^4)N^{-8.122636\times10^{-5}}$	$21.9168 + (7.410873)C^{-6.959190\times10^{-2}}$
	DenseNet	$6829.2802 + (-6.802892 \times 10^3)N^{-1.243441\times10^{-4}}$	$-494.7395 + (5.238394 \times 10^2)C^{-3.616424\times10^{-4}}$
	EfficientNet	$32.6059 + (-6.718342)N^{-2.341518\times10^{-1}}$	$-494.7395 + (5.238394 \times 10^2)C^{-3.616424\times10^{-4}}$

Table 19. 5-way 5-shot power laws for models trained on ImageNet.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$26.7636 + (-7.550182 \times 10^1)N^{-2.862002}$	$26.5106 + (-2.830452 \times 10^1)C^{-2.029323}$
	ResNet	$28.8560 + (-2.420079 \times 10^1)N^{-1.588830}$	$27.5949 + (-8.213628)C^{-8.442404\times10^{-1}}$
	VGG	$31.5332 + (-5.601404)N^{-1.678293\times10^{-1}}$	$71.5770 + (-4.713489 \times 10^1)C^{-1.230203\times10^{-2}}$
	DenseNet	$30.4520 + (-9.944618)N^{-7.982078\times10^{-1}}$	$27.9134 + (-4.312050 \times 10^1)C^{-2.021588}$
	EfficientNet	$4219.8089 + (-4.192289 \times 10^3)N^{-1.496625\times10^{-4}}$	$29.2210 + (-1.029941 \times 10^1)C^{-5.783265\times10^{-1}}$
Bird	CNN	$227.7161 + (-2.006267 \times 10^2)N^{-9.848532\times10^{-5}}$	$27.3531 + (-3.177685)C^{-5.724016\times10^{-1}}$
	ResNet	$36.6348 + (-1.576946 \times 10^1)N^{-8.820985\times10^{-1}}$	$34.8112 + (-1.171667 \times 10^1)C^{-4.973628\times10^{-1}}$
	VGG	$38.3844 + (-1.127407 \times 10^1)N^{-3.699856\times10^{-1}}$	$33.0232 + (-2.182021 \times 10^1)C^{-9.649279\times10^{-1}}$
	DenseNet	$37.6183 + (-1.388570 \times 10^1)N^{-6.885964\times10^{-1}}$	$33.9220 + (-1.757090 \times 10^1)C^{-1.051029}$
	EfficientNet	$39.1082 + (-8.546102)N^{-3.675151\times10^{-1}}$	$35.1290 + (-2.475406 \times 10^1)C^{-9.643434\times10^{-1}}$
Desc. Text.	CNN	$29.0936 + (-2.731010 \times 10^1)N^{-1.942451}$	$28.4514 + (-2.065079 \times 10^1)C^{-1.492969}$
	ResNet	$35.2235 + (-1.342253 \times 10^1)N^{-6.263512\times10^{-1}}$	$31.3815 + (-2.677248)C^{-6.658988\times10^{-1}}$
	VGG	$37.4923 + (-1.043264 \times 10^1)N^{-2.239192\times10^{-1}}$	$30.7430 + (-9.321503)C^{-1.147026}$
	DenseNet	$36.2620 + (-9.864120)N^{-4.632775\times10^{-1}}$	$32.2126 + (-5.831859 \times 10^1)C^{-2.961560}$
	EfficientNet	$39.2167 + (-1.097644 \times 10^1)N^{-1.954921\times10^{-1}}$	$32.4758 + (-9.802830)C^{-5.600419\times10^{-1}}$
Flower	CNN	$34.0602 + (-4.190498 \times 10^2)N^{-4.598250}$	$35.8537 + (-6.566147)C^{-2.545880\times10^{-1}}$
	ResNet	$38.5463 + (-7.489522)N^{-9.621347\times10^{-1}}$	$37.4385 + (-1.107866 \times 10^1)C^{-8.173230\times10^{-1}}$
	VGG	$40.3052 + (-5.370151)N^{-2.503540\times10^{-1}}$	$38.5928 + (-1.005607 \times 10^1)C^{-4.202883\times10^{-1}}$
	DenseNet	$39.2809 + (-9.641306)N^{-9.585369\times10^{-1}}$	$37.8877 + (-4.344661)C^{-9.440823\times10^{-1}}$
	EfficientNet	$39.9229 + (-2.818440)N^{-3.600089\times10^{-1}}$	$38.5655 + (-2.756924 \times 10^1)C^{-1.136270}$
Fungi	CNN	$-1056.0835 + (1.085194 \times 10^3)N^{-1.539402\times10^{-4}}$	$29.4057 + (-9.648819)C^{-6.611459\times10^{-1}}$
	ResNet	$34.4883 + (-2.046771 \times 10^1)N^{-1.275939}$	$37.2085 + (-1.152195 \times 10^1)C^{-2.096094\times10^{-1}}$
	VGG	$37.0654 + (-8.193690)N^{-3.534857\times10^{-1}}$	$33.3767 + (-2.840130 \times 10^1)C^{-1.082893}$
	DenseNet	$40.4596 + (-9.643328)N^{-1.882294\times10^{-1}}$	$13745.0352 + (-1.371529 \times 10^4)C^{-7.987764\times10^{-5}}$
	EfficientNet	$2063.9784 + (-2.031607 \times 10^3)N^{-4.263579\times10^{-4}}$	$34.9430 + (-2.299524 \times 10^1)C^{-7.741862\times10^{-1}}$
Omniglot	CNN	$27.4094 + (1.072140)N^{-2.974458\times10^{-1}}$	$34.3015 + (-8.697018)C^{-6.137529\times10^{-2}}$
	ResNet	$34.7469 + (1.343210)N^{-1.487116\times10^{-1}}$	$34.7681 + (-4.596332 \times 10^1)C^{-2.061814}$
	VGG	$2935.8094 + (-2.902497 \times 10^3)N^{-1.117086\times10^{-4}}$	$34.3181 + (-1.427115 \times 10^1)C^{-9.938190\times10^{-1}}$
	DenseNet	$2209.8185 + (-2.175651 \times 10^3)N^{-2.198611\times10^{-4}}$	$834.4937 + (-7.997191 \times 10^2)C^{-1.720679\times10^{-4}}$
	EfficientNet	$36.3211 + (-2.932900 \times 10^1)N^{-1.540103}$	$35.5593 + (-4.828714 \times 10^1)C^{-1.749749}$
Traffic Sign	CNN	$2136.8259 + (-2.110141 \times 10^3)N^{-1.116872\times10^{-4}}$	$26.8492 + (5.732261 \times 10^{-1})C^{-2.923474\times10^{-1}}$
	ResNet	$32.6637 + (-1.435734 \times 10^{14})N^{-1.792281\times10^1}$	$1447.4866 + (-1.417750 \times 10^3)C^{-4.255031\times10^{-4}}$
	VGG	$33.4101 + (-7.497739)N^{-4.112143\times10^{-1}}$	$29.8873 + (-1.003388 \times 10^1)C^{-1.349309}$
	DenseNet	$34.3264 + (-2.720936 \times 10^1)N^{-1.379141}$	$219.4125 + (-1.872639 \times 10^2)C^{-1.679230\times10^{-4}}$
	EfficientNet	$33.6484 + (-2.725150 \times 10^1)N^{-1.560226}$	$32.4493 + (-1.095418 \times 10^1)C^{-1.053851}$
Average	CNN	$28.6961 + (-3.160969 \times 10^1)N^{-2.865791}$	$28.6617 + (-5.245146)C^{-7.362728\times10^{-1}}$
	ResNet	$34.3660 + (-1.731822 \times 10^1)N^{-1.299510}$	$33.3805 + (-6.942833)C^{-5.353376\times10^{-1}}$
	VGG	$36.4190 + (-7.154206)N^{-2.665333\times10^{-1}}$	$32.3538 + (-1.159193 \times 10^1)C^{-8.329950\times10^{-1}}$
	DenseNet	$36.1568 + (-8.075576)N^{-5.508379\times10^{-1}}$	$34.0510 + (-3.316947)C^{-3.809393\times10^{-1}}$
	EfficientNet	$36.3476 + (-6.379826)N^{-4.299884\times10^{-1}}$	$33.8953 + (-1.791221 \times 10^1)C^{-9.297659\times10^{-1}}$

Table 20. 5-way 5-shot power laws for models trained on Omniglot.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$25.1993 + (9.036498 \times 10^{-1})N^{-2.425596\times10^{-1}}$	$1385.9830 + (-1.360831 \times 10^{3})C^{-1.355002\times10^{-4}}$
	ResNet	$282.1985 + (-2.592942 \times 10^{2})N^{-8.464373\times10^{-5}}$	$23.4267 + (-2.487869)C^{-3.367486\times10^{-1}}$
	VGG	$267.4134 + (-2.443545 \times 10^{2})N^{-1.607903\times10^{-4}}$	$23.4060 + (-4.566062)C^{-8.201303\times10^{-1}}$
	DenseNet	$429.0005 + (-4.067772 \times 10^{2})N^{-1.733047\times10^{-4}}$	$396.4139 + (-3.743232 \times 10^{2})C^{-9.200200\times10^{-5}}$
	EfficientNet	$23.2778 + (-2.840143 \times 10^{13})N^{-1.733557\times10^{1}}$	$3393.4685 + (-3.372058 \times 10^{3})C^{-1.145117\times10^{-4}}$
Bird	CNN	$23.4045 + (-2.448008)N^{-6.026950\times10^{-1}}$	$22.6350 + (2.698433)C^{-5.352016\times10^{-1}}$
	ResNet	$22.8958 + (-3.170261 \times 10^{2})N^{-3.327987}$	$365.4263 + (-3.429467 \times 10^{2})C^{-1.218546\times10^{-4}}$
	VGG	$-403.7119 + (4.270007 \times 10^{2})N^{-3.720575\times10^{-4}}$	$23.0846 + (-2.731116 \times 10^{1})C^{-1.934489}$
	DenseNet	$1224.9067 + (-1.203177 \times 10^{3})N^{-1.879740\times10^{-4}}$	$-737.8247 + (7.605406 \times 10^{2})C^{-1.229306\times10^{-4}}$
	EfficientNet	$22.6270 + (-9.162429 \times 10^{10})N^{-1.420641\times10^{1}}$	$2342.6608 + (-2.321408 \times 10^{3})C^{-1.045756\times10^{-4}}$
Desc. Text.	CNN	$696.5300 + (-6.737683 \times 10^{2})N^{-1.583947\times10^{-4}}$	$22.8826 + (4.741109)C^{-7.810437\times10^{-1}}$
	ResNet	$22.7290 + (-1.713817 \times 10^{1})N^{-1.688829}$	$369.6347 + (-3.477233 \times 10^{2})C^{-3.875342\times10^{-4}}$
	VGG	$23.0556 + (-2.103817)N^{-4.484163\times10^{-1}}$	$-552.1323 + (5.750994 \times 10^{2})C^{-2.074553\times10^{-4}}$
	DenseNet	$22.3384 + (-9.821382)N^{-1.977748}$	$22.1907 + (-4.181652 \times 10^{1})C^{-2.129080}$
	EfficientNet	$-785.3889 + (8.082542 \times 10^{2})N^{-1.267848\times10^{-4}}$	$3076.9299 + (-3.056033 \times 10^{3})C^{-1.210715\times10^{-4}}$
Flower	CNN	$2066.9753 + (-2.041051 \times 10^{3})N^{-7.575991\times10^{-5}}$	$25.0335 + (5.551881)C^{-3.488706\times10^{-1}}$
	ResNet	$25.6338 + (-1.739551 \times 10^{1})N^{-1.262726}$	$331.9138 + (-3.090230 \times 10^{2})C^{-1.353580\times10^{-3}}$
	VGG	$25.5299 + (-2.251183)N^{-8.134813\times10^{-1}}$	$25.5280 + (-2.640766)C^{-6.873529\times10^{-1}}$
	DenseNet	$24.9311 + (-1.694829)N^{-8.553403\times10^{-1}}$	$24.4726 + (-3.649337 \times 10^{1})C^{-1.863362}$
	EfficientNet	$25.4747 + (-4.737602 \times 10^{13})N^{-1.710718\times10^{1}}$	$3733.8817 + (-3.710480 \times 10^{3})C^{-1.301629\times10^{-4}}$
Fungi	CNN	$23.9259 + (3.804613 \times 10^{1})N^{-2.473682}$	$23.9122 + (3.280921)C^{-6.999613\times10^{-1}}$
	ResNet	$24.3489 + (-6.007774 \times 10^{14})N^{-1.856983\times10^{1}}$	$2079.1298 + (-2.056254 \times 10^{3})C^{-1.455136\times10^{-4}}$
	VGG	$24.5605 + (-2.489155 \times 10^{-1})N^{-1.984203\times10^{-1}}$	$317.3053 + (-2.934143 \times 10^{2})C^{-1.248983\times10^{-4}}$
	DenseNet	$24.8407 + (-3.059707 \times 10^{-1})N^{2.858475\times10^{-1}}$	$2169.2063 + (-2.145886 \times 10^{3})C^{-1.073874\times10^{-4}}$
	EfficientNet	$24.0391 + (-1.121077 \times 10^{3})N^{-3.771453}$	$-552.9249 + (5.774104 \times 10^{2})C^{-1.920872\times10^{-4}}$
Traffic Sign	CNN	$26.0716 + (1.928818 \times 10^{1})N^{-1.744868}$	$570.0226 + (-5.442199 \times 10^{2})C^{-3.037806\times10^{-4}}$
	ResNet	$3729.1947 + (-3.702282 \times 10^{3})N^{1.627317\times10^{-4}}$	$758.4348 + (-7.369106 \times 10^{2})C^{-1.150207\times10^{-3}}$
	VGG	$23.3169 + (1.566210)N^{-1.204322\times10^{-1}}$	$24.9069 + (-4.757143 \times 10^{1})C^{-1.629285}$
	DenseNet	$2332.8958 + (-2.309408 \times 10^{3})N^{-7.637684\times10^{-5}}$	$24.7645 + (-1.595877)C^{-1.087328\times10^{-1}}$
	EfficientNet	$24.2463 + (2.021908)N^{-3.401310\times10^{-1}}$	$2543.5592 + (-2.520975 \times 10^{3})C^{-1.791128\times10^{-4}}$
Average	CNN	$24.7250 + (2.285094 \times 10^{1})N^{-3.073321}$	$24.6974 + (3.357999)C^{-8.396358\times10^{-1}}$
	ResNet	$23.8260 + (-2.566703 \times 10^{14})N^{-1.851554\times10^{1}}$	$2256.6182 + (-2.234380 \times 10^{3})C^{-1.486913\times10^{-4}}$
	VGG	$23.8208 + (-3.025174 \times 10^{-1})N^{-5.779324\times10^{-1}}$	$23.9116 + (-3.119690)C^{-9.212053\times10^{-1}}$
	DenseNet	$323.7769 + (-3.006626 \times 10^{2})N^{-2.359043\times10^{-4}}$	$261.3383 + (-2.384780 \times 10^{2})C^{-3.416952\times10^{-4}}$
	EfficientNet	$23.7973 + (-2.035644 \times 10^{13})N^{-1.717737\times10^{1}}$	$2182.4927 + (-2.160154 \times 10^{3})C^{-1.406164\times10^{-4}}$

Table 21. 5-way 5-shot power laws for models trained on Quickdraw.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Quickdraw	CNN	$-535.6703 + (5.618870 \times 10^2)N^{-1.427195\times10^{-4}}$	$2138.5021 + (-2.113721 \times 10^3)C^{-1.334645\times10^{-4}}$
	ResNet	$24.4460 + (-2.317429 \times 10^{13})N^{-1.770256\times10^1}$	$24.6520 + (-1.197931 \times 10^3)C^{-2.941889}$
	EfficientNet	$24.8788 + (-1.197454 \times 10^{13})N^{-1.722997\times10^1}$	$24.6901 + (-4.859561 \times 10^1)C^{-1.491496}$
Bird	CNN	$-1017.1821 + (1.040554 \times 10^3)N^{-1.058313\times10^{-4}}$	$1499.6765 + (-1.477451 \times 10^3)C^{-1.272158\times10^{-4}}$
	ResNet	$213.7498 + (-1.908605 \times 10^2)N^{-1.632290\times10^{-4}}$	$23.2483 + (-2.740925 \times 10^{19})C^{-1.815517\times10^1}$
	EfficientNet	$22.5301 + (5.263514)N^{-1.396521}$	$22.6913 + (-2.386181 \times 10^{19})C^{-1.819083\times10^1}$
Desc. Text.	CNN	$-502.6509 + (5.268460 \times 10^2)N^{-1.821075\times10^{-4}}$	$3366.7587 + (-3.344276 \times 10^3)C^{-9.817851\times10^{-5}}$
	ResNet	$24.0090 + (-1.182822 \times 10^{12})N^{-1.647314\times10^1}$	$23.9467 + (-3.667660 \times 10^{15})C^{-1.489007\times10^1}$
	EfficientNet	$23.5030 + (3.249048 \times 10^1)N^{-2.813994}$	$23.5137 + (-1.816036 \times 10^3)C^{-3.409820}$
Flower	CNN	$-484.5320 + (5.119992 \times 10^2)N^{-1.174421\times10^{-4}}$	$2691.3230 + (-2.665233 \times 10^3)C^{-9.890083\times10^{-5}}$
	ResNet	$-646.9176 + (6.737624 \times 10^2)N^{-1.375925\times10^{-4}}$	$26.3884 + (-9.557852 \times 10^8)C^{-8.259501}$
	EfficientNet	$27.0486 + (-1.418310 \times 10^{-34})N^{1.661147\times10^1}$	$27.3389 + (-2.331565 \times 10^1)C^{-1.016222}$
Fungi	CNN	$24.7383 + (6.431768)N^{-2.433589}$	$101.4406 + (-7.707841 \times 10^1)C^{-2.782688\times10^{-4}}$
	ResNet	$-1159.4641 + (1.184509 \times 10^3)N^{-1.168086\times10^{-4}}$	$24.7711 + (-2.333025)C^{-3.933015\times10^{-1}}$
	EfficientNet	$24.6629 + (2.562788 \times 10^1)N^{-2.331371}$	$25.3183 + (-4.928283 \times 10^{19})C^{-1.795817\times10^1}$
Omniglot	CNN	$-494.2591 + (5.318705 \times 10^2)N^{-1.881613\times10^{-4}}$	$1071.6678 + (-1.035138 \times 10^3)C^{-1.915309\times10^{-4}}$
	ResNet	$-2862.3062 + (2.902456 \times 10^3)N^{-1.127026\times10^{-4}}$	$42.1405 + (-6.571713)C^{-1.639711\times10^{-1}}$
	EfficientNet	$2090.8972 + (-2.055264 \times 10^3)N^{-2.196221\times10^{-4}}$	$38.6901 + (-3.303391 \times 10^1)C^{-9.134228\times10^{-1}}$
Traffic Sign	CNN	$-2087.5120 + (2.114249 \times 10^3)N^{-1.740476\times10^{-4}}$	$2131.1663 + (-2.106505 \times 10^3)C^{-1.103199\times10^{-4}}$
	ResNet	$24.4400 + (2.895295)N^{-1.009365}$	$24.8789 + (-6.227238 \times 10^{17})C^{-1.623051\times10^1}$
	EfficientNet	$188.8555 + (-1.638603 \times 10^2)N^{1.142500\times10^{-4}}$	$1126.4569 + (-1.102609 \times 10^3)C^{-1.591129\times10^{-4}}$
Average	CNN	$-1122.5950 + (1.149783 \times 10^3)N^{-9.972996\times10^{-5}}$	$1951.7047 + (-1.925828 \times 10^3)C^{-1.123074\times10^{-4}}$
	ResNet	$-951.5501 + (9.785033 \times 10^2)N^{-1.100165\times10^{-4}}$	$26.6043 + (-6.892365 \times 10^2)C^{-2.773483}$
	EfficientNet	$369.7339 + (-3.434554 \times 10^2)N^{-1.116894\times10^{-4}}$	$26.5544 + (-9.612072 \times 10^1)C^{-1.806139}$

Table 22. 5-way 5-shot power laws for models trained on Traffic Sign.

Target	Model	$Acc(N)$ (training data)	$Acc(C)$ (training classes)
Aircraft	CNN	$26.6380 + (-2.363741)N^{-3.466945\times10^{-1}}$	$431.7916 + (-4.069537\times10^{2})C^{-1.184380\times10^{-4}}$
	ResNet	$23.6223 + (4.620080\times10^{13})N^{-1.819177\times10^{1}}$	$23.1706 + (-5.151756\times10^{2})C^{-2.429009}$
	VGG	$23.5184 + (-2.472198)N^{-7.895254\times10^{-1}}$	$22.5477 + (-9.455361\times10^{4})C^{-4.834902}$
	DenseNet	$2100.1511 + (-2.077975\times10^{3})N^{-1.694825\times10^{-4}}$	$22.6898 + (-1.875872\times10^{1})C^{-1.596342}$
	EfficientNet	$23.2166 + (-6.962095)N^{-1.291386}$	$22.6654 + (-9.215854\times10^{4})C^{-4.499645}$
Bird	CNN	$-644.9423 + (6.709212\times10^{2})N^{-1.966290\times10^{-4}}$	$26.9592 + (-3.420423)C^{-2.194002\times10^{-1}}$
	ResNet	$2332.1925 + (-2.307072\times10^{3})N^{2.324778\times10^{-4}}$	$6422.1425 + (-6.403212\times10^{3})C^{-1.761421\times10^{-4}}$
	VGG	$22.7661 + (4.633320)N^{-2.195719\times10^{-1}}$	$2958.9477 + (-2.938160\times10^{3})C^{-3.498420\times10^{-4}}$
	DenseNet	$24.4296 + (1.794696)N^{-3.901559\times10^{-1}}$	$27.0130 + (-5.942419)C^{-1.967704\times10^{-1}}$
	EfficientNet	$3170.9005 + (-3.147126\times10^{3})N^{1.237709\times10^{-4}}$	$3543.2994 + (-3.524222\times10^{3})C^{-2.571639\times10^{-4}}$
Desc. Text.	CNN	$24.0449 + (6.147811\times10^{-2})N^{-5.591198\times10^{-1}}$	$24.0327 + (-4.107694\times10^{18})C^{-1.748799\times10^{1}}$
	ResNet	$22.5093 + (3.204275)N^{-9.331176\times10^{-1}}$	$1520.6364 + (-1.499996\times10^{3})C^{-3.104055\times10^{-4}}$
	VGG	$-1221.4643 + (1.245133\times10^{3})N^{-1.485905\times10^{-4}}$	$2341.4092 + (-2.320326\times10^{3})C^{-1.946092\times10^{-4}}$
	DenseNet	$462.7831 + (-4.395589\times10^{2})N^{-1.314647\times10^{-4}}$	$1356.8753 + (-1.335184\times10^{3})C^{-2.587611\times10^{-4}}$
	EfficientNet	$22.2172 + (4.027027\times10^{8})N^{-1.199519\times10^{1}}$	$23.1258 + (-7.394190)C^{-4.784568\times10^{-1}}$
Flower	CNN	$31.5912 + (-1.723717\times10^{14})N^{-1.842053\times10^{1}}$	$31.8696 + (-5.180918\times10^{1})C^{-1.512799}$
	ResNet	$223.0439 + (-1.924589\times10^{2})N^{4.018394\times10^{-3}}$	$5241.9298 + (-5.219321\times10^{3})C^{-2.611745\times10^{-4}}$
	VGG	$-4117.2105 + (4.149084\times10^{3})N^{-1.562124\times10^{-4}}$	$30.8849 + (-1.137235\times10^{1})C^{-4.864300\times10^{-1}}$
	DenseNet	$-1256.4816 + (1.287586\times10^{3})N^{-1.929927\times10^{-4}}$	$34.5508 + (-1.048465\times10^{1})C^{-1.892917\times10^{-1}}$
	EfficientNet	$1625.9028 + (-1.599187\times10^{3})N^{1.096650\times10^{-4}}$	$34.2346 + (-1.447838\times10^{1})C^{-1.451210\times10^{-1}}$
Fungi	CNN	$1313.2010 + (-1.286858\times10^{3})N^{-1.149577\times10^{-4}}$	$26.5017 + (4.051022\times10^{2})C^{-2.755194}$
	ResNet	$22.1263 + (4.879869)N^{-5.254080\times10^{-1}}$	$23.4218 + (-2.313126\times10^{1})C^{-8.732697}$
	VGG	$20.6850 + (8.359943)N^{-1.920485\times10^{-1}}$	$1553.5094 + (-1.530839\times10^{3})C^{-4.141569\times10^{-4}}$
	DenseNet	$-2158.1716 + (2.184576\times10^{3})N^{-2.138293\times10^{-4}}$	$3386.3913 + (-3.363359\times10^{3})C^{-1.255170\times10^{-4}}$
	EfficientNet	$2191.9449 + (-2.168137\times10^{3})N^{1.121088\times10^{-4}}$	$23.2142 + (-9.434387\times10^{1})C^{-1.759574}$
Omniglot	CNN	$40.0084 + (-3.075476)N^{-2.803200\times10^{-1}}$	$38.2400 + (7.588299\times10^{17})C^{-1.680449\times10^{1}}$
	ResNet	$8175.6127 + (-8.145819\times10^{3})N^{-9.046507\times10^{-5}}$	$32.8174 + (-4.459182\times10^{2})C^{-1.979514}$
	VGG	$3934.9420 + (-3.898503\times10^{3})N^{1.215743\times10^{-4}}$	$33.8117 + (-4.380637\times10^{1})C^{-1.098876\times10^{1}}$
	DenseNet	$3613.9880 + (-3.575973\times10^{3})N^{1.817029\times10^{-4}}$	$35.2133 + (-2.527034\times10^{20})C^{-1.846312\times10^{1}}$
	EfficientNet	$34.8820 + (-8.669142)N^{-4.495056\times10^{-1}}$	$10877.7027 + (-1.085879\times10^{4})C^{-2.678306\times10^{-4}}$
Average	CNN	$29.0962 + (-1.004641)N^{-3.452423\times10^{-1}}$	$1075.3669 + (-1.047360\times10^{3})C^{-1.191391\times10^{-4}}$
	ResNet	$25.2612 + (4.395870)N^{-1.005276}$	$26.8368 + (-1.257110\times10^{1})C^{-5.886205\times10^{-1}}$
	VGG	$23.1461 + (5.347964)N^{-1.116673\times10^{-1}}$	$28.4770 + (-1.031648\times10^{1})C^{-4.363656\times10^{-1}}$
	DenseNet	$26.3721 + (1.703572)N^{-2.583005\times10^{-1}}$	$27.9203 + (-4.877519)C^{-3.259682\times10^{-1}}$
	EfficientNet	$651.4805 + (-6.267455\times10^{2})N^{-9.612340\times10^{-5}}$	$29.2078 + (-1.108146\times10^{1})C^{-2.160407\times10^{-1}}$